

Achievable Rates for Channels with Deletions and Insertions

Ramji Venkataramanan

Yale University

ramji.venkataramanan@yale.edu

Sekhar Tatikonda*

Yale University

sekhar.tatikonda@yale.edu

Kannan Ramchandran

University of California, Berkeley

kannanr@eecs.berkeley.edu

January 13, 2013

Abstract

This paper considers a binary channel with deletions and insertions, where each input bit is transformed in one of the following ways: it is deleted with probability d , or an extra bit is added after it with probability i , or it is transmitted unmodified with probability $1 - d - i$. A computable lower bound on the capacity of this channel is derived. The transformation of the input sequence by the channel may be viewed in terms of runs as follows: some runs of the input sequence get shorter/longer, some runs get deleted, and some new runs are added. It is difficult for the decoder to synchronize the channel output sequence to the transmitted codeword mainly due to deleted runs and new inserted runs. We consider a decoder that decodes the positions of the deleted and inserted runs in addition to the transmitted codeword. Analyzing the performance of such a decoder leads to a computable lower bound on the capacity. The bounds proposed in this paper provide the first characterization of achievable rates for channels with general insertions, and for channels with both deletions and insertions. For the special cases of deletion channels and duplication channels where previous results exist, our rates are very close to the best-known capacity lower bounds.

1 Introduction

Consider a binary input channel where for each bit (denoted x), the output is generated in one of the following ways:

- The bit is deleted with probability d ,
- An extra bit is inserted after x with probability i . The extra bit is equal to x (a *duplication*) with probability α , and equal to $1 - x$ (a *complementary insertion*) with probability $1 - \alpha$,
- No deletions or insertions occur, and the output is x with probability $1 - d - i$.

*This work was partially supported by NSF Grant CCF-1017744. It will be presented in part at the 2011 IEEE International Symposium on Information Theory.

The channel acts independently on each bit. We refer to this channel as the deletion+insertion channel with parameters (d, i, α) . If the input to the channel is a sequence of n bits, the length of the output sequence will be close to $n(1 + i - d)$ for large n due to the law of large numbers.

Channels with synchronization errors can be used to model timing mismatch in communication systems. Channels with deletions and insertions also occur in magnetic recording [1]. The problem of synchronization also appears in file backup and file sharing [2, 3], where distributed nodes may have different versions of the same file which differ by a small number of edits. The edits may include deletions, insertions, and substitutions. The minimum communication rate required to synchronize the remote sources is closely related to the capacity of an associated synchronization channel. This connection is discussed at the end of this paper.

The above model with $i = 0$ corresponds to the deletion channel, which has been studied in several recent papers, e.g., [4–11]. When $d = 0$, we obtain the insertion channel with parameters (i, α) . The insertion channel with $\alpha = 1$ is the sticky channel [12], where all insertions are duplications.

In this paper, we obtain lower bounds on the capacity of the deletion+insertion channel. Our starting point is the result of Dobrushin [13] for general synchronization channels which states that the capacity is given by the maximum of the mutual information per bit between the input and output sequences. There are two challenges to computing the capacity through this characterization. The first is evaluating the mutual information, which is a difficult task because of the memory inherent in the joint distribution of the input and output sequences. The second challenge is to optimize the mutual information over all input distributions.

In this work, we choose the input distribution to be the class of first-order Markov processes and focus on the problem of evaluating the mutual information. It is known that first-order Markov input distributions yield good capacity lower bounds for the deletion channel [4, 5] and the sticky channel [12], both special cases of the deletion+insertion channel. This suggests they are likely to perform well on the deletion+insertion channel as well. But our approach is quite general, and can be extended to obtain capacity lower bounds with other input distributions such as higher order Markov distributions, and distributions with independent run-lengths.

For a synchronization channel, it is useful to think of the input and output sequences in terms of runs of symbols rather than individual symbols. (The runs of a binary sequence are its alternating blocks of contiguous zeros and ones.) If there were a one-to-one correspondence between the runs of the input sequence \underline{X} and those of the output sequence \underline{Y} , we could write the conditional distribution $P(\underline{Y}|\underline{X})$ as a product distribution of run-length transformations; computing the mutual information would then be straightforward. Unfortunately, such a correspondence is not possible since deletions can lead to some runs being lost, and insertions to new runs being inserted.

The main idea of the paper is to use auxiliary sequences which indicate the positions (in the output sequence) where *runs* were deleted and inserted. Consider a decoder that first decodes the auxiliary sequences, and then the transmitted codeword. Conditioned on the knowledge of the auxiliary sequences, the mutual information between the input and output sequences is a single-letter quantity which can be calculated easily. However, decoding the auxiliary sequences in addition to the codeword is sub-optimal and incurs a rate penalty, which we upper bound. The challenge is to define auxiliary sequences that lead to analytical lower bounds on the capacity with minimal rate penalty.

To gain insight, we first consider the special cases of the insertion channel and the deletion channel separately. The insertion channel with parameters (i, α) introduces approximately ni insertions in a sufficiently long input sequence of length n . A fraction nearly α of these insertions are duplications, and the rest are

complementary insertions. Note that new runs can only be introduced by complementary insertions. We consider a decoder that first decodes the positions of the complementary insertions. The decoder can then flip the bits at these positions to obtain a one-to-one correspondence between input and output runs.

For the deletion channel, we use a decoder that first decodes an auxiliary sequence whose symbols indicate the number of runs deleted between each pair of adjacent bits in the output sequence. Augmenting the output sequence with the positions of deleted runs results in a one-to-one correspondence between input and output runs. For the deletion+insertion channel, we consider a decoder that decodes both auxiliary sequences described above.

The main contributions of the paper are the following:

1. Theorems 1 and 2 together provide the first characterization of achievable rates for the general insertion channel. For the special case of the ‘sticky’ channel ($\alpha = 1$, i.e., only duplications), the rates of Theorem 2 are very close to the near-optimal lower bound given in [12].
2. Theorem 4 provides the first characterization of achievable rates for the deletion+insertion channel. For the special case of the deletion channel ($i = 0$), these rates are close to the best-known lower bounds given in [5].
3. Our approach provides a general framework to compute the capacity of channels with synchronization errors, and suggests several directions to obtain sharper capacity bounds. For example, results on the structure of optimal input distributions for these channels (in the spirit of [10, 11]) could be combined with our approach to improve the lower bounds. One could also obtain upper bounds on the capacity by assuming that the auxiliary sequences are available ‘for free’ at the decoder, as done in [4] for the deletion channel.

For clarity, we only consider the binary deletion+insertion channel in this paper. The results presented here can be extended to channels with any finite alphabet.

1.1 Related Work

Dobrushin’s capacity characterization was used in [6] to obtain bounds on the deletion capacity. The ‘jigsaw’ decoding approach of [5] is interpreted in [6] in terms of an auxiliary sequence that associates each run of the output sequence with one or more runs of the input codeword. Using this, the mutual information is decomposed such that one part of it represents the rate achieved by the jigsaw decoder, and the other part represents the rate loss due to the jigsaw decoder. This rate loss is hard to compute, and is estimated via simulation for a few values of the deletion probability in [6].

Our approach is motivated by the observation that synchronizing the output with the transmitted sequence is difficult mainly due to runs being completely deleted and new runs being inserted. Accordingly, the auxiliary sequences we use indicate the positions of these runs in the output sequence, leading to a mutual information decomposition that is quite different from [6]. The approach of [5, 6] provides the best-known lower bounds on the deletion capacity, but is specific to channels with only deletions and duplications. Our techniques are general and apply to channels with both deletions and insertions, while giving bounds very close to the best-known ones for deletion and duplication channels.

Dobrushin’s capacity characterization was also used in [11] to estimate the deletion capacity and the structure of the optimal input distribution for small values of deletion probability. In [7], a genie-aided decoder

with access to the locations of deleted runs was used to upper bound the deletion capacity using an equivalent discrete memoryless channel (DMC). In [9], bounds on the deletion capacity were obtained by considering a decoder equipped with side-information specifying the number of output bits corresponding to successive blocks of L input bits, for any positive integer L . This new channel is equivalent to a DMC with an input alphabet of size 2^L , whose capacity can be numerically computed using the Blahut-Arimoto algorithm (for as large a value of L as computationally feasible). The upper bound in [9] is the best known for a wide range of deletion probabilities, but the lower bound is weaker than that of [5] and the one proposed here. Finally, we note that a different channel model with bit flips and synchronization errors was studied in [14, 15]. In this model, an insertion is defined as an input bit being replaced by two random bits. We have only mentioned the papers that are closely related to the results of this work. The reader is referred to [8] for an exhaustive list of references on synchronization channels.

After laying down the formal definitions and technical machinery in Section 2, we describe two coding schemes in Section 3 which give intuition about our bounding techniques. In Section 4, we consider the insertion channel ($d = 0$) and derive two lower bounds on its capacity. For this channel, previous bounds exist only for the special case of sticky channels ($\alpha = 1$) [12]. We derive a lower bound on the capacity of the deletion channel ($i = 0$) in Section 5 and compare it with the best known lower bound. In Section 6, we combine the ideas of Sections 4 and 5 to obtain a lower bound for the deletion+insertion channel. Section 7 concludes the paper with a discussion of open questions.

2 Preliminaries

Notation: \mathbb{N}_0 denotes the set of non-negative integers, and \mathbb{N} the set of natural numbers. $h(\cdot)$ is the binary entropy function, and $\mathbf{1}_{\mathcal{A}}$ is the indicator function of the set \mathcal{A} . For any $0 < \alpha \leq 1$, $\bar{\alpha} \triangleq 1 - \alpha$. Logarithms are with base 2, and entropy is measured in bits. We use uppercase letters to denote random variables, bold-face letters for random processes, and superscript notation to denote random vectors. Thus the channel input sequence of length n is denoted $X^n \triangleq (X_1, \dots, X_n)$. The corresponding output sequence at the decoder has length M_n (a random variable determined by the channel realization), and is denoted Y^{M_n} . For brevity, we sometimes use underlined notation for random vectors when we do not need to be explicit about their length. Thus $\underline{X} \triangleq X^n = (X_1, X_2, \dots, X_n)$, and $\underline{Y} \triangleq Y^{M_n} = (Y_1, \dots, Y_{M_n})$.

The communication over the channel is characterized by three random processes defined over the same probability space: the input process $\mathbf{X} = \{X_n\}_{n \geq 1}$, the output process $\mathbf{Y} = \{Y_n\}_{n \geq 1}$, and $\mathbf{M} = \{M_n\}_{n \geq 1}$, where M_n is the number of output symbols corresponding to the first n input symbols. If the underlying probability space is (Ω, \mathcal{F}, P) , each realization $\omega \in \Omega$ determines the sample paths $\mathbf{X}(\omega) = \{X_n(\omega)\}_{n \geq 1}$, $\mathbf{Y}(\omega) = \{Y_n(\omega)\}_{n \geq 1}$, and $\mathbf{M}(\omega) = \{M_n(\omega)\}_{n \geq 1}$.

Definition 2.1. An $(n, 2^{nR})$ code with block length n and rate R consists of

1. An encoder mapping $e : \{1, \dots, 2^{nR}\} \rightarrow \{0, 1\}^n$, and
2. A decoder mapping $g : \Sigma \rightarrow \{1, \dots, 2^{nR}\}$ where Σ is $\cup_{k=0}^n \{0, 1\}^k$ for the deletion channel, $\cup_{k=n}^{2n} \{0, 1\}^k$ for the insertion channel, and $\cup_{k=0}^{2n} \{0, 1\}^k$ for the deletion+insertion channel.

Assuming the message W is drawn uniformly on the set $\{1, \dots, 2^{nR}\}$, the probability of error of a $(n, 2^{nR})$

code is

$$P_{e,n} = \frac{1}{2^{nR}} \sum_{l=1}^{2^{nR}} \Pr(g(Y^{M_n}) \neq l | W = l)$$

A rate R is achievable if there exists a sequence of $(n, 2^{nR})$ codes such that $P_{e,n} \rightarrow 0$ as $n \rightarrow \infty$. The supremum of all achievable rates is the capacity C . The following characterization of capacity follows from a result proved for a general class of synchronization channels by Dobrushin [13].

Fact 1. *Let $C_n = \max_{P_{X^n}} \frac{1}{n} I(X^n; Y^{M_n})$. Then $C \triangleq \lim_{n \rightarrow \infty} C_n$ exists, and is equal to the capacity of the deletion+insertion channel.*

Proof. Dobrushin proved the following general result in [13]. Consider a channel with \mathcal{X} and \mathcal{Y} denoting the alphabets of possible symbols at the input and output, respectively. For each input symbol in \mathcal{X} , the output belongs to $\bar{\mathcal{Y}}$, the set of all finite sequences of elements of \mathcal{Y} , including the empty sequence. The channel is memoryless and is specified by the stochastic matrix $\{P(\bar{y}|x), \bar{y} \in \bar{\mathcal{Y}}, x \in \mathcal{X}\}$. Also assume that for each input symbol x , the length of the (possibly empty) output sequence has non-zero expected value. Then $\lim_{n \rightarrow \infty} C_n$ exists, and is equal to the capacity of the channel.

The deletion+insertion channel is a special case of the above model with $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, and the length of the output corresponding to any input symbol has a maximum value of two and expected value equal to $(1 - d + i)$, which is non-zero for all $d < 1$. Hence the claim is a direct consequence of Dobrushin's result. \square

In this paper, we fix the input process to be the class of binary symmetric first-order Markov processes and focus on evaluating the mutual information. This will give us a lower bound on the capacity. The input process $\mathbf{X} = \{X_n\}_{n \geq 1}$ is characterized by the following distribution for all n :

$$P(X_1, \dots, X_n) = P(X_1) \prod_{j=2}^n P(X_j | X_{j-1}),$$

with

$$P(X_1 = 0) = P(X_1 = 1) = 0.5, \quad P(X_j = 1 | X_{j-1} = 1) = P(X_j = 0 | X_{j-1} = 0) = \gamma, \quad j \geq 1. \quad (1)$$

A binary sequence may be represented by a sequence of positive integers representing the lengths of its runs, and the value of the first bit (to indicate whether the first run has zeros or ones). For example, the sequence 0001100000 can be represented as $(3, 2, 5)$ if we know that the first bit is 0. The value of the first bit of \mathbf{X} can be communicated to the decoder with vanishing rate, and we will assume this has been done at the outset. Hence, denoting the length of the j th run of \mathbf{X} by L_j^X we have the following equivalence: $\mathbf{X} \leftrightarrow (L_1^X, L_2^X, \dots)$. For a first-order Markov binary source of (1), the run-lengths are independent and geometrically distributed, i.e.,

$$\Pr(L_j^X = r) = \gamma^{r-1}(1 - \gamma), \quad r = 1, 2, \dots \quad (2)$$

The average length of a run in \mathbf{X} is $\frac{1}{1-\gamma}$, so the number of runs in a sequence of length n is close to $n(1 - \gamma)$ for large n . Our bounding techniques aim to establish a one-to-one correspondence between input runs and output runs. The independence of run-lengths of \mathbf{X} enables us to obtain analytical bounds on the capacity.

We denote by $I_P(X^n; Y^{M_n})$, $H_P(X^n)$, $H_P(X^n|Y^{M_n})$ the mutual information and entropies computed with the channel input sequence X^n distributed as in (1). For all n , we have

$$C_n = \max_{P_{X^n}} \frac{1}{n} I(X^n; Y^{M_n}) > \frac{1}{n} I_P(X^n; Y^{M_n}). \quad (3)$$

Therefore

$$C > \liminf_{n \rightarrow \infty} \frac{1}{n} I_P(X^n; Y^{M_n}) = h(\gamma) - \limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n|Y^{M_n}). \quad (4)$$

We will derive upper bounds on $\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n|Y^{M_n})$ and use it in (4) to obtain a lower bound on the capacity.

2.1 Technical Lemmas

To formally prove our results, we will use a framework similar to [6]. The notion of uniform integrability will play an important role, and we list the relevant definitions and technical lemmas below.

Definition 2.2. [16] *A family of random variables $\{Z_n\}_{n \geq 1}$ is uniformly integrable if*

$$\lim_{a \rightarrow \infty} \sup_n \mathbf{E}[|Z_n| \mathbf{1}_{\{|Z_n| \geq a\}}] = 0.$$

Lemma 2.1. [16] *A family of random variables $\{Z_n\}_{n \geq 1}$ is uniformly integrable if and only if:*

1. $\sup_n \mathbf{E}[|Z_n|] < \infty$, and
2. For any $\epsilon > 0$, there exists some $\delta > 0$ such that for all n and any event \mathcal{A} with $\Pr(\mathcal{A}) < \delta$, we have $\mathbf{E}[|Z_n| \mathbf{1}_{\mathcal{A}}] < \epsilon$.

Let $\text{Supp}(W|Z)$ denote the random variable whose value is the size of the support of the conditional distribution of W given Z .

Lemma 2.2. [6, Lemma 4] *Let $\{W_n, Z_n\}_{n \geq 1}$ be a sequence of pairs of discrete random variables with $\text{Supp}(W_n|Z_n) \leq c^n$ for some constant $c \geq 1$. Then $\sup_n \mathbf{E}\left[\left(\frac{1}{n} \log \Pr(W_n|Z_n)\right)^2\right] < \infty$. In particular, the sequence $\left\{-\frac{1}{n} \log \Pr(W_n|Z_n)\right\}_{n \geq 1}$ is uniformly integrable.*

Lemma 2.3. [16] *Suppose that $\{Z_n : n \geq 1\}$ is a sequence of random variables that converges to Z in probability. Then the following are equivalent.*

1. $\{Z_n : n \geq 1\}$ is uniformly integrable.
2. $\mathbf{E}[|Z_n|] < \infty$ for all n , and $\mathbf{E}[|Z_n|] \rightarrow \mathbf{E}[|Z|]$ as $n \rightarrow \infty$.

Lemma 2.4. *Let $\mathbf{Z} = \{Z_n\}_{n \geq 1}$ be a process for the asymptotic equipartition property (AEP) holds, i.e.,*

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \Pr(Z_1, \dots, Z_n) \rightarrow H(Z) \quad \text{a.s.}$$

where $H(Z)$ is the (finite) entropy rate of the process \mathbf{Z} . Let $\{M_n\}_{n \geq 1}$ be a sequence of positive integer valued random variables defined on the same probability space as the Z_n 's, and suppose that $\lim_{n \rightarrow \infty} \frac{M_n}{n} = x$ almost

surely for some constant x . Then

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \Pr(Z_1, \dots, Z_{M_n}) = H(Z)x \quad a.s.$$

Proof. Fix any $\epsilon > 0$. Since $\lim_{n \rightarrow \infty} \frac{M_n}{n} = x$, there exists an $L(\epsilon)$ such that for all $n > L(\epsilon)$, we have almost surely

$$a(n, \epsilon) \triangleq \lceil n(x - \epsilon) \rceil \leq M_n \leq \lceil n(x + \epsilon) \rceil \triangleq b(n, \epsilon).$$

It follows that for all $n > L(\epsilon)$,

$$-\frac{1}{n} \log \Pr(Z_1, \dots, Z_{M_n}) \geq -\frac{1}{n} \log \Pr(Z_1, \dots, Z_{a(n, \epsilon)}) = -\frac{a(n, \epsilon)}{n} \cdot \frac{\log \Pr(Z_1, \dots, Z_{a(n, \epsilon)})}{a(n, \epsilon)}. \quad (5)$$

Hence,

$$\begin{aligned} \liminf_{n \rightarrow \infty} -\frac{1}{n} \log \Pr(Z_1, \dots, Z_{M_n}) &\geq \liminf_{n \rightarrow \infty} -\frac{a(n, \epsilon)}{n} \cdot \frac{\log \Pr(Z_1, \dots, Z_{a(n, \epsilon)})}{a(n, \epsilon)} \\ &= \lim_{n \rightarrow \infty} \frac{a(n, \epsilon)}{n} \cdot \lim_{n \rightarrow \infty} \frac{-\log \Pr(Z_1, \dots, Z_{a(n, \epsilon)})}{a(n, \epsilon)} \\ &= (x - \epsilon)H(Z). \end{aligned} \quad (6)$$

Similarly, one can show that

$$\limsup_{n \rightarrow \infty} -\frac{1}{n} \log \Pr(Z_1, \dots, Z_{M_n}) \leq (x + \epsilon)H(Z). \quad (7)$$

Since $\epsilon > 0$ is arbitrary, combining (6) and (7), we get the result of the lemma. \square

3 Coding Schemes

In this section, we describe coding schemes to give intuition about the auxiliary sequences used to obtain the bounds. The discussion is informal; the capacity bounds are rigorously proved in the following sections using a different technique: the auxiliary sequences are used to directly bound the limiting behavior of $\frac{1}{n}H(X^n|Y^{M_n})$ using information-theoretic inequalities and elementary tools from analysis.

3.1 Insertion Channel

Consider the insertion channel with parameters (i, α) . For $0 < \alpha < 1$, the inserted bits may create new runs, so we cannot associate each run of \underline{Y} with a run in \underline{X} . For example, let

$$\underline{X} = 000111000 \quad \text{and} \quad \underline{Y} = 00 \textit{101110000} \textit{0}, \quad (8)$$

where the inserted bits are indicated in large italics. There is one duplication (in the third run), and two complementary insertions (in the first and second runs). While a duplication never introduces a new run, a complementary insertion introduces a new run, except when it occurs at the end of a run of \underline{X} (e.g., the 0 inserted at the end of the second run in (8)). For any input-pair (X^n, Y^{M_n}) , define an auxiliary sequence $T^{M_n} = (T_1, \dots, T_{M_n})$ where $T_j = 1$ if Y_j is a *complementary* insertion, and $T_j = 0$ otherwise.

The sequence T^{M_n} indicates the positions of the complementary insertions in Y^{M_n} . In the example of (8), $T^{M_n} = (0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0)$.

Consider the following coding scheme. Construct a codebook of 2^{nR} codewords of length n , each chosen independently according to the first-order Markov distribution(1). Let X^n denote the transmitted codeword, and Y^{M_n} the channel output. From Y^{M_n} , the decoder decodes (using joint typicality) the positions of the complementary insertions, in addition to the input sequence. The joint distribution of these sequences is determined by the input distribution (1) and the channel parameters (i, α) .

Such a decoder is sub-optimal since the complementary insertion pattern T^{M_n} is not unique given an input-output pair (X^n, Y^{M_n}) . This is discussed in Section 4. The maximum rate achievable by this decoder is obtained by analyzing the probability of error. Assuming all sequences satisfy the asymptotic equipartition property [17], we have for sufficiently large n

$$\Pr(\text{error}) \leq 2^{n(R+H(T^{M_n}|X^n))} \cdot 2^{-nI(X^n T^{M_n}; Y^{M_n})}. \quad (9)$$

The second term above is the probability that (X^n, T^{M_n}, Y^{M_n}) are jointly typical when Y^{M_n} is picked independently from (X^n, T^{M_n}) . The first term is obtained by taking a union bound over all the codewords and all the typical complementary insertion patterns for each codeword. Hence the probability of error goes to zero if

$$R < \frac{1}{n} (I(X^n T^{M_n}; Y^{M_n}) - H(T^{M_n}|X^n)) = \frac{1}{n} (H(X^n) - H(T^{M_n}|Y^{M_n}) - H(X^n|T^{M_n}, Y^{M_n})). \quad (10)$$

We obtain a lower bound on the capacity in Section 4.2 by obtaining good single-letter bounds on the limiting behavior of both $\frac{1}{n}H(T^{M_n}|Y^{M_n})$ and $\frac{1}{n}H(X^n|T^{M_n}, Y^{M_n})$.

3.2 Deletion Channel

For the deletion channel with deletion probability d , consider the following pair of input and output sequences $\underline{X} = 000111000$, $\underline{Y} = 0010$. For this pair, we can associate each run of \underline{Y} uniquely with a run in \underline{X} . Therefore, we can write

$$P(\underline{Y} = 0010|\underline{X} = 000111000) = P(L_1^Y = 2|L_1^X = 3)P(L_2^Y = 1|L_2^X = 3)P(L_3^Y = 1|L_3^X = 3)$$

where L_j^X, L_j^Y denote the lengths of the j th runs of X and Y , respectively. We observe that if no runs in \underline{X} are completely deleted, then the conditional distribution of \underline{Y} given \underline{X} may be written as a product distribution of run-length transformations:

$$P(\underline{Y}|\underline{X}) = P(L_1^Y|L_1^X)P(L_2^Y|L_2^X)P(L_3^Y|L_3^X)\dots \quad (11)$$

where for all runs j , $P(L_j^Y = s|L_j^X = r) = \binom{r}{s}d^{r-s}(1-d)^s$ for $1 \leq s \leq r$. In general, we do have runs of \underline{X} that are completely deleted. For example, if $\underline{X} = 000111000$ and $\underline{Y} = 000$, we cannot associate the single run in \underline{Y} uniquely with a run in \underline{X} .

For any input-output pair (X^n, Y^{M_n}) , define an auxiliary sequence $S^{M_n+1} = (S_1, S_2, \dots, S_{M_n+1})$, where $S_j \in \mathbb{N}_0$ is the number of runs *completely* deleted in X^n between the bits corresponding to Y_{j-1} and Y_j . (S_1 is the number of runs deleted before the first output symbol Y_1 , and S_{M_n+1} is the number of runs deleted after

the last output symbol Y_{M_n} .) For example, if $\underline{X} = 00 \underline{0111000}$ and the bits shown in italics were deleted to give $\underline{Y} = 000$, then $\underline{S} = (0, 0, 1, 0)$. On the other hand, if the last six bits were all deleted, i.e., $\underline{X} = 000 \underline{111000}$, then $\underline{S} = (0, 0, 0, 2)$. Thus \underline{S} is not uniquely determined given $(\underline{X}, \underline{Y})$. The auxiliary sequence \underline{S} enables us to augment \underline{Y} with the positions of missing runs. As will be explained in Section 5, the runs of this augmented output sequence are in one-to-one correspondence with the runs of the input sequence.

Consider the following coding scheme. Construct a codebook of 2^{nR} codewords of length n , each chosen independently according to (1). The decoder receives Y^{M_n} , and decodes (using joint typicality) both the auxiliary sequence and the input sequence. Such a decoder is sub-optimal since the auxiliary sequence S^{M_n+1} is not unique given a codeword X^n and the output Y^{M_n} . Assuming all sequences satisfy the asymptotic equipartition property, we have for sufficiently large n

$$\Pr(\text{error}) \leq 2^{n(R+H(S^{M_n+1}|X^n))} \cdot 2^{-nI(X^n S^{M_n+1}; Y^{M_n})}. \quad (12)$$

The second term above is the probability that $(X^n, S^{M_n+1}, Y^{M_n})$ are jointly typical when Y^{M_n} is picked independently from (X^n, S^{M_n+1}) . The first term is obtained by taking a union bound over all the codewords and all the typical auxiliary sequences for each codeword. Hence the probability of error goes to zero if

$$R < \frac{1}{n} (I(X^n S^{M_n+1}; Y^{M_n}) - H(S^{M_n+1}|X^n)) = \frac{1}{n} (H(X^n) - H(S^{M_n+1}|Y^{M_n}) - H(X^n|S^{M_n+1}, Y^{M_n})). \quad (13)$$

In Section 5, we show that the above expression converges as $n \rightarrow \infty$, and obtain an analytical expression for the limit.

For the deletion+insertion channel, we use both the auxiliary sequences, T^{M_n} and S^{M_n+1} . The decoder decodes both these sequences in addition to the codeword X^n , and the maximum achievable rate is given by

$$R < \frac{1}{n} (H(X^n) - H(T^{M_n}, S^{M_n+1}|Y^{M_n}) - H(X^n|S^{M_n+1}, T^{M_n}, Y^{M_n})). \quad (14)$$

We obtain a lower bound on the capacity of the deletion+insertion channel in Section 6 by analyzing the limiting behavior of (14).

4 Insertion Channel

In this channel, an extra bit may be inserted after each bit of \underline{X} with probability $i \in (0, 1)$. When a bit is inserted after X_j , the inserted bit is equal to X_j (a duplication) with probability α , and equal to \bar{X}_j (a complementary insertion) with probability $1 - \alpha$. When $\alpha = 1$, we have only duplications - this is the sticky channel studied in [12]. In this case, we can associate each run of \underline{Y} with a unique run in \underline{X} , which leads to a computable single-letter characterization of the best achievable rates with a first-order Markov distribution. We derive two lower bounds on the capacity of the insertion channel, each using a different auxiliary sequence.

4.1 Lower Bound 1

For any input-pair (X^n, Y^{M_n}) , define an auxiliary sequence $I^{M_n} = (I_1, \dots, I_{M_n})$ where $I_j = 1$ if Y_j is an inserted bit, and $I_j = 0$ otherwise. The sequence I^{M_n} indicates the positions of all the inserted bits in Y^{M_n} ,

and is not unique for a given (X^n, Y^{M_n}) . Using I^{M_n} , we can decompose $H_P(X^n|Y^{M_n})$ as

$$H_P(X^n|Y^{M_n}) = H_P(X^n, I^{M_n}|Y^{M_n}) - H_P(I^{M_n}|X^n, Y^{M_n}) = H_P(I^{M_n}|Y^{M_n}) - H_P(I^{M_n}|X^n, Y^{M_n})$$

since $H(X^n|Y^{M_n}, I^{M_n}) = 0$. Therefore,

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n|Y^{M_n}) &= \limsup_{n \rightarrow \infty} \frac{1}{n} (H_P(I^{M_n}|Y^{M_n}) - H_P(I^{M_n}|X^n, Y^{M_n})) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} H_P(I^{M_n}|Y^{M_n}) - \liminf_{n \rightarrow \infty} \frac{1}{n} H_P(I^{M_n}|X^n, Y^{M_n}). \end{aligned} \quad (15)$$

We will derive an upper bound on $\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(I^{M_n}|Y^{M_n})$, and a lower bound on $\liminf_{n \rightarrow \infty} \frac{1}{n} H_P(I^{M_n}|X^n, Y^{M_n})$. Using this in (15), we get an upper bound for $\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n|Y^{M_n})$, which can then be used in (4).

Proposition 4.1. *The process $\{\mathbf{I}, \mathbf{Y}\} \triangleq \{(I_1, Y_1), (I_2, Y_2), \dots\}$ is a second-order Markov process characterized by the following joint distribution for all $m \in \mathbb{N}$:*

$$P(I^m, Y^m) = P(I_1, Y_1)P(I_2, Y_2|I_1, Y_1) \prod_{j=3}^m P(I_j, Y_j|I_{j-1}, Y_{j-1}, Y_{j-2})$$

where for $x, y \in \{0, 1\}$ and $j \geq 3$:

$$\begin{aligned} P(I_j = 1, Y_j = y|I_{j-1} = 0, Y_{j-1} = y, Y_{j-2} = x) &= i\alpha, \quad P(I_j = 1, Y_j = \bar{y}|I_{j-1} = 0, Y_{j-1} = y, Y_{j-2} = x) = i\bar{\alpha} \\ P(I_j = 0, Y_j = y|I_{j-1} = 0, Y_{j-1} = y, Y_{j-2} = x) &= \bar{i}\gamma, \quad P(I_j = 0, Y_j = \bar{y}|I_{j-1} = 0, Y_{j-1} = y, Y_{j-2} = x) = \bar{i}\bar{\gamma} \\ P(I_j = 0, Y_j = x|I_{j-1} = 1, Y_{j-1} = y, Y_{j-2} = x) &= \gamma, \quad P(I_j = 0, Y_j = \bar{x}|I_{j-1} = 1, Y_{j-1} = y, Y_{j-2} = x) = \bar{\gamma} \end{aligned} \quad (16)$$

Proof. We need to show that for all $j \geq 3$, the following Markov relation holds: $(I_j, Y_j) - (I_{j-1}, Y_{j-1}, Y_{j-2}) - (I^{j-2}, Y^{j-2})$. First consider $P(I_j, Y_j|I_{j-1} = 0, Y_{j-1} = y, I^{j-2}, Y^{j-2})$. Since $I_{j-1} = 0$, Y_{j-1} is the most recent input bit (say X_a) before Y_j . $P(I_j = 0, Y_j = y|I_{j-1} = 0, Y_{j-1} = y, I^{j-2}, Y^{j-2})$ is the probability that the following independent events both occur: 1) the input bit X_{a+1} equals X_a , and 2) there was no insertion after input bit X_a . Since the insertion process is i.i.d and is independent of the first-order Markov input process \mathbf{X} , we have

$$P(I_j = 0, Y_j = y|I_{j-1} = 0, Y_{j-1} = y, I^{j-2}, Y^{j-2}) = (1 - i)\gamma.$$

Similarly, we obtain

$$\begin{aligned} P(I_j = 0, Y_j = \bar{y}|I_{j-1} = 0, Y_{j-1} = y, I^{j-2}, Y^{j-2}) &= (1 - i)(1 - \gamma), \\ P(I_j = 1, Y_j = y|I_{j-1} = 0, Y_{j-1} = y, I^{j-2}, Y^{j-2}) &= i\alpha, \\ P(I_j = 1, Y_j = \bar{y}|I_{j-1} = 0, Y_{j-1} = y, I^{j-2}, Y^{j-2}) &= i(1 - \alpha). \end{aligned}$$

Next consider $P(I_j, Y_j|I_{j-1} = 1, Y_{j-1} = y, Y_{j-2} = x, I^{j-2}, Y^{j-2})$. Since $I_{j-1} = 1$, Y_{j-2} is the most recent input bit (say, X_a) before Y_j . Also note that Y_j is the input bit X_{a+1} since Y_{j-1} is an insertion. (At most one insertion can occur after each input bit.) Hence $P(I_j = 0, Y_j = x|I_{j-1} = 1, Y_{j-1} = y, Y_{j-2} = x, I^{j-2}, Y^{j-2})$ is

just the probability that $X_{a+1} = X_a$, which is equal to γ . Similarly,

$$P(I_j = 0, Y_j = \bar{x} | I_{j-1} = 1, Y_{j-1} = y, Y_{j-2} = x, I^{j-2}, Y^{j-2}) = 1 - \gamma.$$

□

Remark: Proposition 4.1 implies that the process $\{\mathbf{I}, \mathbf{Y}\}$ can be characterized as a Markov chain with state at time j given by (I_j, Y_j, Y_{j-1}) . This is an aperiodic, irreducible Markov chain. Hence a stationary distribution π exists, which can be verified to be

$$\begin{aligned} \pi(I_j = 1, Y_j = y, Y_{j-1} = y) &= \frac{i\alpha}{2(1+i)}, & \pi(I_j = 1, Y_j = \bar{y}, Y_{j-1} = y) &= \frac{i\bar{\alpha}}{2(1+i)}, \\ \pi(I_j = 0, Y_j = y, Y_{j-1} = y) &= \frac{\bar{i}\gamma + i\alpha\gamma + i\bar{\alpha}\bar{\gamma}}{2(1+i)}, & \pi(I_j = 0, Y_j = \bar{y}, Y_{j-1} = y) &= \frac{\bar{i}\bar{\gamma} + i\alpha\bar{\gamma} + i\bar{\alpha}\gamma}{2(1+i)} \end{aligned} \quad (17)$$

for $y \in \{0, 1\}$.

Lemma 4.1. $\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(I^{M_n} | Y^{M_n}) = (1+i) \limsup_{m \rightarrow \infty} \frac{1}{m} H_P(I^m | Y^m)$.

Proof. See Appendix A.1. □

Lemma 4.2. $\limsup_{m \rightarrow \infty} \frac{1}{m} H_P(I^m | Y^m) \leq \lim_{j \rightarrow \infty} H_P(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2})$, and

$$\lim_{j \rightarrow \infty} H_P(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2}) = \frac{(i\alpha + \bar{i}\gamma)}{1+i} h\left(\frac{i\alpha}{i\alpha + \bar{i}\gamma}\right) + \frac{(i\bar{\alpha} + \bar{i}\bar{\gamma})}{1+i} h\left(\frac{i\bar{\alpha}}{i\bar{\alpha} + \bar{i}\bar{\gamma}}\right). \quad (18)$$

Proof. See Appendix A.2. □

Next, we focus on $H(I^{M_n} | Y^{M_n}, X^n)$, which is the uncertainty in the positions of the insertions given both the channel input and output sequences. For example, given input $\underline{X} = 0\ 0\ 0\ 1\ 0$, and output $\underline{Y} = 0\ 0\ 0\ 1\ 1\ 0$, we know that there is either a complementary insertion after the third bit of \underline{X} or a duplication after the fourth bit; so there is uncertainty in the values of I_4 and I_5 (one of them is zero, and the other is one.). But there is no uncertainty in I_1, I_2, I_3, I_6 , which are all zero. We use this intuition to obtain a lower bound on the limiting behavior of $\frac{1}{n} H(I^{M_n} | Y^{M_n}, X^n)$.

Lemma 4.3. $\liminf_{n \rightarrow \infty} \frac{1}{n} H_P(I^{M_n} | Y^{M_n}, X^n) = \liminf_{n \rightarrow \infty} \frac{1}{n} H_P(I^{n(1+i)} | Y^{n(1+i)}, X^n)$.

Proof. The proof of this lemma is similar to that of Lemma 4.1, and is omitted. □

Lemma 4.4. $\liminf_{n \rightarrow \infty} \frac{1}{n} H_P(I^{n(1+i)} | Y^{n(1+i)}, X^n) \geq \bar{\gamma}^2 i (\bar{\alpha} + \bar{i}\bar{\gamma}) h\left(\frac{\bar{\alpha}}{\bar{\alpha} + \bar{i}\bar{\gamma}}\right)$.

Proof. See Appendix A.3. □

Theorem 1. (LB 1) *The capacity of the insertion channel with parameters (i, α) can be lower bounded as*

$$C(i, \alpha) \geq \max_{0 < \gamma < 1} h(\gamma) - (i\alpha + \bar{i}\gamma) h\left(\frac{i\alpha}{i\alpha + \bar{i}\gamma}\right) - (i\bar{\alpha} + \bar{i}\bar{\gamma}) h\left(\frac{i\bar{\alpha}}{i\bar{\alpha} + \bar{i}\bar{\gamma}}\right) + \bar{\gamma}^2 i (\bar{\alpha} + \bar{i}\bar{\gamma}) h\left(\frac{\bar{\alpha}}{\bar{\alpha} + \bar{i}\bar{\gamma}}\right).$$

Proof. The result is obtained by using Lemmas 4.1, 4.2, 4.3, and 4.4 in (15), and substituting the resulting bound in (4). We optimize the lower bound by maximizing over the Markov parameter $\gamma \in (0, 1)$. □

4.2 Lower Bound 2

For any input-pair (X^n, Y^{M_n}) , define an auxiliary sequence $T^{M_n} = (T_1, \dots, T_{M_n})$ where $T_j = 1$ if Y_j is a *complementary* insertion, and $T_j = 0$ otherwise. The sequence T^{M_n} indicates the positions of the complementary insertions in Y^{M_n} . Note that T^{M_n} is different from the sequence I^{M_n} , which indicates the positions of *all* the insertions. Using T^{M_n} , we can decompose $H_P(X^n|Y^{M_n})$ as

$$\begin{aligned} H_P(X^n|Y^{M_n}) &= H_P(X^n, T^{M_n}|Y^{M_n}) - H_P(T^{M_n}|X^n, Y^{M_n}) \\ &= H_P(T^{M_n}|Y^{M_n}) + H_P(X^n|T^{M_n}, Y^{M_n}) - H_P(T^{M_n}|X^n, Y^{M_n}) \\ &\stackrel{(a)}{\leq} H_P(T^{M_n}|Y^{M_n}) + H_P(X^n|\tilde{Y}^{M_n}) - H_P(T^{M_n}|X^n, Y^{M_n}), \end{aligned} \quad (19)$$

where \tilde{Y}^{M_n} is the sequence obtained from (T^{M_n}, Y^{M_n}) by flipping Y_j whenever $T_j = 1$, for $1 \leq j \leq M_n$. (a) holds in (19) because \tilde{Y}^{M_n} is a function of (T^{M_n}, Y^{M_n}) , and hence $H_P(X^n|T^{M_n}, Y^{M_n}) \leq H(X^n|\tilde{Y}^{M_n})$. Therefore, we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n|Y^{M_n}) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \left(H_P(T^{M_n}|Y^{M_n}) + H_P(X^n|\tilde{Y}^{M_n}) - H_P(T^{M_n}|X^n, Y^{M_n}) \right). \quad (20)$$

We will show that $\lim_{n \rightarrow \infty} \frac{1}{n} H(X^n|\tilde{Y}^{M_n})$ exists, and therefore (20) becomes

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n|Y^{M_n}) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} (H_P(T^{M_n}|Y^{M_n}) - H_P(T^{M_n}|X^n, Y^{M_n})) + \lim_{n \rightarrow \infty} \frac{1}{n} H_P(X^n|\tilde{Y}^{M_n}) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} H_P(T^{M_n}|Y^{M_n}) - \liminf_{n \rightarrow \infty} \frac{1}{n} H_P(T^{M_n}|X^n, Y^{M_n}) + \lim_{n \rightarrow \infty} \frac{1}{n} H_P(X^n|\tilde{Y}^{M_n}) \end{aligned} \quad (21)$$

We will use (21) in (4) to obtain a lower bound on the insertion capacity.

Lemma 4.5. $\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(T^{M_n}|Y^{M_n}) = (1+i) \limsup_{m \rightarrow \infty} \frac{1}{m} H_P(T^m|Y^m)$.

Proof. The proof of this lemma is identical to that of Lemma 4.1, and can be obtained T^{M_n} replacing I^{M_n} . \square

Lemma 4.6. $\limsup_{m \rightarrow \infty} \frac{1}{m} H_P(T^m|Y^m) \leq \lim_{j \rightarrow \infty} H_P(T_j|T_{j-1}, Y_j, Y_{j-1})$, and

$$\lim_{j \rightarrow \infty} H_P(T_j|T_{j-1}, Y_j, Y_{j-1}) = \frac{(1-\gamma + \gamma i \bar{\alpha})}{(1+i)} h \left(\frac{i \bar{\alpha}}{1-\gamma + \gamma i \bar{\alpha}} \right). \quad (22)$$

Proof. See Appendix A.4. \square

We now determine the limiting behavior of $\frac{1}{n} H(X^n|\tilde{Y}^{M_n})$. Recall that \tilde{Y}^{M_n} is obtained by flipping the complementary insertions in Y^{M_n} . In other words, \tilde{Y}^{M_n} has insertions in the same locations as Y^{M_n} , but the insertions are all duplications. Hence \tilde{Y}^{M_n} has the same number of runs as X^n . Recall from Section 2 that we can represent both binary sequences in terms of their run-lengths as

$$X^n \leftrightarrow (L_1^X, \dots, L_{R_n}^X), \quad \tilde{Y}^{M_n} \leftrightarrow (L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}}),$$

where R_n , the number of runs in X^n (and \tilde{Y}^{M_n}) is a random variable. Therefore, for all n , we have

$$H_P(X^n|\tilde{Y}^{M_n}) = H_P(L_1^X, \dots, L_{R_n}^X | L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}}). \quad (23)$$

Proposition 4.2. *The process $\{\mathbf{L}^{\mathbf{X}}, \mathbf{L}^{\tilde{\mathbf{Y}}}\} \triangleq \{(L_1^{\mathbf{X}}, L_1^{\tilde{\mathbf{Y}}}), (L_2^{\mathbf{X}}, L_2^{\tilde{\mathbf{Y}}}), \dots\}$ is an i.i.d process characterized by the following joint distribution for all $j \geq 1$:*

$$P(L_j^{\mathbf{X}} = r, L_j^{\tilde{\mathbf{Y}}} = s) = \gamma^{r-1}(1-\gamma) \cdot \binom{r}{s-r} i^{s-r}(1-i)^{2r-s}, \quad r = 1, 2, \dots, r \leq s \leq 2r.$$

Proof. Since \mathbf{X} is a Markov process, $\{L_j^{\mathbf{X}}\}_{j \geq 1}$ are independent with

$$P(L_j^{\mathbf{X}} = r) = \gamma^{r-1}(1-\gamma), \quad r = 1, 2, \dots$$

\tilde{Y}^{M_n} is generated from X^n by independently duplicating each bit with probability i . Hence $L_j^{\tilde{\mathbf{Y}}}$ can be thought of being obtained by passing a run of length $L_j^{\mathbf{X}}$ through a discrete memoryless channel with transition probability

$$P(L_j^{\tilde{\mathbf{Y}}} = s | L_j^{\mathbf{X}} = r) = \binom{r}{s-r} i^{s-r}(1-i)^{2r-s}, \quad r \leq s \leq 2r.$$

□

Lemma 4.7. $\lim_{n \rightarrow \infty} \frac{1}{n} H_P(X^n | \tilde{Y}^{M_n}) = (1-\gamma) H_P(L_1^{\mathbf{X}} | L_1^{\tilde{\mathbf{Y}}})$, where the joint distribution of $(L_1^{\mathbf{X}}, L_1^{\tilde{\mathbf{Y}}})$ is given by Proposition 4.2.

Proof. See Appendix A.5. □

Finally, we need to analyze $H(T^{M_n} | Y^{M_n}, X^n)$, the uncertainty in the positions of the complementary insertions given both the channel input and output sequences. For example, given input $\underline{X} = 0 \ 0 \ 0 \ 1 \ 0$, and output $\underline{Y} = 0 \ 0 \ 0 \ 1 \ 1 \ 0$, we know that there is either a complementary insertion after the third bit of \underline{X} or a duplication after the fourth bit; so there is uncertainty in the value of T_4 . There is no uncertainty in T_1, T_2, T_3, T_5, T_6 , which are all zero. We use this intuition to obtain a lower bound on the limiting behavior of $\frac{1}{n} H(T^{M_n} | Y^{M_n}, X^n)$.

Lemma 4.8. $\liminf_{n \rightarrow \infty} \frac{1}{n} H_P(T^{M_n} | Y^{M_n}, X^n) = \liminf_{n \rightarrow \infty} \frac{1}{n} H_P(T^{n(1+i)} | Y^{n(1+i)}, X^n)$.

Proof. The proof of this lemma is similar to that of Lemma 4.1, and is omitted. □

Lemma 4.9. $\liminf_{n \rightarrow \infty} \frac{1}{n} H_P(T^{n(1+i)} | Y^{n(1+i)}, X^n) \geq \bar{\gamma}^2 i(\bar{\alpha} + \bar{i}\alpha) h\left(\frac{\bar{\alpha}}{\bar{\alpha} + \bar{i}\alpha}\right)$.

Proof. See Appendix A.6. □

Theorem 2. (LB 2) *The capacity of the insertion channel with parameters (i, α) can be lower bounded as*

$$C(i, \alpha) \geq \max_{0 < \gamma < 1} h(\gamma) - (\bar{\gamma} + \gamma i \bar{\alpha}) h\left(\frac{i \bar{\alpha}}{\bar{\gamma} + \gamma i \bar{\alpha}}\right) - \bar{\gamma} H(L_{X1} | L_{\tilde{Y}1}) + \bar{\gamma}^2 i(\bar{\alpha} + \bar{i}\alpha) h\left(\frac{\bar{\alpha}}{\bar{\alpha} + \bar{i}\alpha}\right)$$

where $H(L_{X1} | L_{\tilde{Y}1})$ is computed using the joint distribution given in Proposition 4.2.

Proof. The result is a direct consequence of using Lemmas 4.5 - 4.9 in (21), and substituting the resulting upper bound for $\limsup_{n \rightarrow \infty} \frac{1}{n} H(X^n | Y^{M_n})$ in (4). □

Figure 1 compares Lower bound 1 with Lower bound 2 for different values of i with α fixed at 0.8. We observe that LB 2 is generally a better bound than LB 1, except when i is large. For large i , it is more efficient

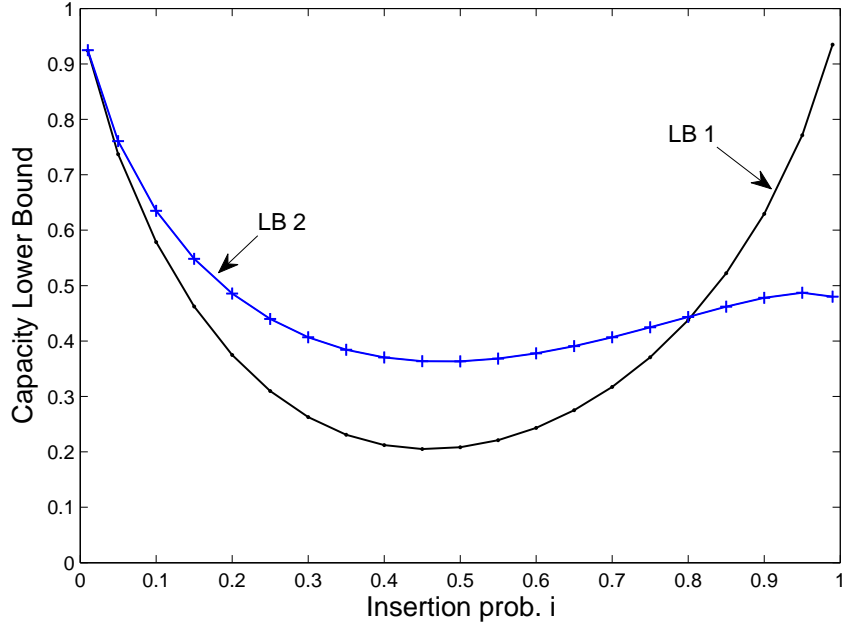


Figure 1: Comparison of the two lower bounds for different values of i with $\alpha = 0.8$

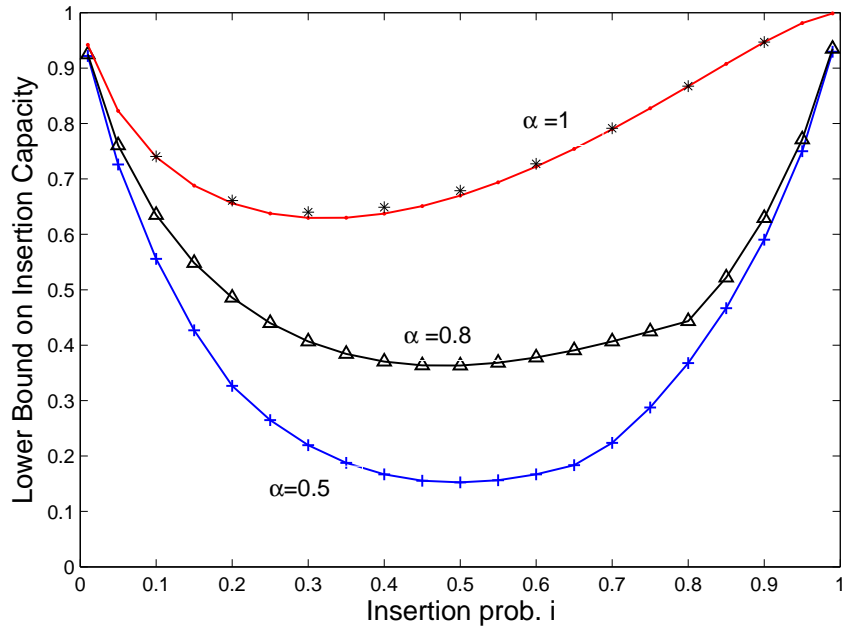


Figure 2: Lower bound $\max\{LB1, LB2\}$ on the insertion capacity $C(i, \alpha)$. For $\alpha = 1$, the lower bound of [12] is shown using '*'.

to decode the positions of all the insertions (since i is large) rather than just the complementary insertions. Specifically, comparing Lemmas 4.2 and 4.6,

$$\lim_{j \rightarrow \infty} H(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2}) \leq \lim_{j \rightarrow \infty} H(T_j | T_{j-1}, Y_j, Y_{j-1})$$

for large values of i . Combining the bounds of Theorems 1 and 2, we observe that $\max\{LB\ 1, LB\ 2\}$ is a lower bound to the insertion capacity. This is plotted in Figure 2 for various values of i , for $\alpha = 0.5, 0.8, 1$. For $\alpha = 1$, the bound is very close to the near-optimal lower bound in [12]. The gap occurs because we have used a Markov input distribution instead of the numerically optimized input distribution in [12].

5 Deletion Channel

In this channel, each input bit is deleted with probability d , or retained with probability $1 - d$. For any input-output pair (X^n, Y^{M_n}) , define the auxiliary sequence S^{M_n+1} , where $S_j \in \mathbb{N}_0$ is the number of runs *completely* deleted in X^n between the bits corresponding to Y_{j-1} and Y_j . (S_1 is the number of runs deleted before the first output symbol Y_1 , and S_{M_n+1} is the number of runs deleted after the last output symbol Y_{M_n} .) Examples of S^{M_n} for the input-output pair ($\underline{X} = 000111000$, $\underline{Y} = 000$) were given in Section 3.2.

The auxiliary sequence \underline{S} enables us to augment \underline{Y} with the positions of missing runs. Consider $\underline{X} = 000111000$. If the decoder was given $\underline{Y} = 000$ and $\underline{S} = (0, 0, 0, 2)$, it can form the augmented sequence $\underline{Y}' = 000 - -$, where a $-$ denotes a missing run, or equivalently a ‘run of length 0’ in \underline{Y} . With the “ $-$ ” markers indicating deleted runs, we can associate each run of the augmented sequence \underline{Y}' uniquely with a run in \underline{X} . Denote by $L_1^{Y'}, L_2^{Y'}, \dots$ the run-lengths of the augmented sequence \underline{Y}' , where $L_j^{Y'} = 0$ if the run is a $-$. Then we have

$$P(\underline{X}, \underline{Y}') = P(L_1^X)P(L_1^{Y'} | L_1^X) \cdot P(L_2^X)P(L_2^{Y'} | L_2^X) \dots \quad (24)$$

where $\forall j$:

$$\begin{aligned} P(L_j^X = r) &= \gamma^{r-1}(1 - \gamma), \quad r = 1, 2, \dots \\ P(L_j^{Y'} = s | L_j^X = r) &= \binom{r}{s} d^{r-s} (1 - d)^s, \quad 0 \leq s \leq r. \end{aligned} \quad (25)$$

Using the auxiliary sequence S^{M_n+1} , we can decompose $H_P(X^n | Y^{M_n})$ as

$$H_P(X^n | Y^{M_n}) = H_P(X^n, S^{M_n+1} | Y^{M_n}) - H_P(S^{M_n+1} | X^n, Y^{M_n}). \quad (26)$$

We therefore have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n | Y^{M_n}) \leq \limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n, S^{M_n+1} | Y^{M_n}) \quad (27)$$

We will show that $\lim_{n \rightarrow \infty} \frac{1}{n} H_P(X^n, S^{M_n+1} | Y^{M_n})$ exists, and obtain an analytical expression for this limit. Using this in (4), we obtain a lower bound on the deletion capacity. We remark that it has been shown in [6] that for any input distribution with independent runs, $\lim_{n \rightarrow \infty} \frac{1}{n} H(X^n | Y^{M_n})$ exists for the deletion channel. Hence the lim sup on the left hand side of (27) is actually a limit.

Proposition 5.1. *The process $\mathbf{Y} = \{Y_1, Y_2, \dots\}$ is a first-order Markov process characterized by the following*

joint distribution for all $m \in \mathbb{N}$.

$$P(Y^m) = P(Y_1) \prod_{j=2}^m P(Y_j | Y_{j-1})$$

where for $y \in \{0, 1\}$

$$P(Y_j = y) = 0.5, \quad P(Y_j = y | Y_{j-1} = y) = 1 - P(Y_j = \bar{y} | Y_{j-1} = y) = \frac{\gamma + d - 2\gamma d}{1 + d - 2\gamma d}. \quad (28)$$

Proof. The proof of this lemma can be found in [8].

Proposition 5.2. *The process $\{\mathbf{S}, \mathbf{Y}\} \triangleq \{(S_1, Y_1), (S_2, Y_2), \dots\}$ is a first-order Markov process characterized by the following joint distribution for all $m \in \mathbb{N}$:*

$$P(S^m, Y^m) = P(Y_1, S_1) \prod_{j=2}^m P(Y_j, S_j | Y_{j-1}),$$

where for $y \in \{0, 1\}$ and $j \geq 2$:

$$P(Y_j = y, S_j = k | Y_{j-1} = y) = \begin{cases} \frac{\gamma(1-d)}{(1-\gamma d)}, & k = 0 \\ \frac{(1-d)(1-\gamma)}{(1-\gamma d)^2} \left(\frac{d(1-\gamma)}{1-\gamma d} \right)^k, & k = 1, 3, \dots \\ 0, & \text{otherwise} \end{cases} \quad (29)$$

$$P(Y_j = \bar{y}, S_j = k | Y_{j-1} = y) = \begin{cases} \frac{(1-d)(1-\gamma)}{(1-\gamma d)^2} \left(\frac{d(1-\gamma)}{1-\gamma d} \right)^k, & k = 0, 2, \dots \\ 0, & \text{otherwise} \end{cases} \quad (30)$$

Proof. We need to show that

$$P(Y_j = y, S_j = k | Y_{j-1} = y_{j-1}, S_{j-1} = s_{j-1}, Y_{j-2} = y_{j-2}, S_{j-2} = s_{j-2}, \dots) = P(Y_j = y, S_j = k | Y_{j-1} = y_{j-1}),$$

for all $y, y_{j-1}, y_{j-2}, \dots \in \{0, 1\}$ and $k, s_{j-1}, \dots \in \mathbb{N}_0$.

Let the output symbols $Y_j, Y_{j-1}, Y_{j-2}, \dots$ correspond to input symbols $X_{a_j}, X_{a_{j-1}}, X_{a_{j-2}}, \dots$ for some positive integers $a_j > a_{j-1} > a_{j-2} > \dots$. S_{j-1} is the number of runs between the input symbols $X_{a_{j-2}}$ and $X_{a_{j-1}}$, not counting the runs containing $X_{a_{j-2}}$ and $X_{a_{j-1}}$. Similarly, S_{j-2} is the number of runs between the input symbols $X_{a_{j-3}}$ and $X_{a_{j-2}}$, not counting the runs containing $X_{a_{j-3}}$ and $X_{a_{j-2}}$ etc.

First consider the case where $Y_j = Y_{j-1} = y$. When $Y_j = X_{a_j} = y$ and $Y_{j-1} = X_{a_{j-1}} = y$, note that S_j , the number of completely deleted runs between $X_{a_{j-1}}$ and X_{a_j} , is either zero or an odd number. We have

$$P(Y_j = y, S_j = 0 | Y_{j-1} = y, S_{j-1} = s_{j-1}, Y_{j-2} = y_{j-2}, S_{j-2} = s_{j-2}, \dots) \stackrel{(a)}{=} \sum_{m=1}^{\infty} \gamma^m (1-\gamma) (1-d^m) = \frac{\gamma(1-d)}{(1-\gamma d)} \quad (31)$$

where (a) is obtained as follows. $\gamma^m (1-\gamma)$ is the probability that the input run containing $X_{a_{j-1}}$ contains m bits after a_{j-1} , and $(1-d^m)$ is the probability that at least one of them is not deleted. This needs to hold for some $m \geq 1$ in order to have $S_j = 0$ and $Y_j = Y_{j-1} = y$. By reasoning similar to the above, we have for

$k = 1, 3, 5, \dots$:

$$\begin{aligned}
& P(Y_j = y, S_j = k | Y_{j-1} = y, S_{j-1} = s_{j-1}, Y_{j-2} = y_{j-2}, S_{j-2} = s_{j-2}, \dots) \\
& \stackrel{(b)}{=} \left(\sum_{m=0}^{\infty} \gamma^m (1-\gamma) d^m \right) \left(\sum_{m=1}^{\infty} \gamma^{m-1} (1-\gamma) d^m \right)^k \left(\sum_{m=1}^{\infty} \gamma^{m-1} (1-\gamma) (1-d^m) \right) \\
& = \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2} \left[\frac{d(1-\gamma)}{(1-\gamma d)} \right]^k
\end{aligned} \tag{32}$$

where the first term in (b) is the probability that the remainder of the run containing $X_{a_{j-1}}$ is completely deleted, the second term is the probability that the next k runs are deleted, and the last term is the probability that the subsequent run is *not* completely deleted.

When $Y_j = y$ and $Y_{j-1} = \bar{y}$, the number of deleted runs S_j is either zero or an even number. For $k = 0, 2, 4, \dots$ we have

$$\begin{aligned}
& P(Y_j = y, S_j = k | Y_{j-1} = \bar{y}, S_{j-1} = s_{j-1}, Y_{j-2} = y_{j-2}, S_{j-2} = s_{j-2}, \dots) \\
& \stackrel{(c)}{=} \left(\sum_{m=1}^{\infty} \gamma^m (1-\gamma) d^m \right) \left(\sum_{m=1}^{\infty} \gamma^{m-1} (1-\gamma) d^m \right)^k \left(\sum_{m=0}^{\infty} \gamma^{m-1} (1-\gamma) (1-d^m) \right) \\
& = \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2} \left[\frac{d(1-\gamma)}{(1-\gamma d)} \right]^k.
\end{aligned} \tag{33}$$

In the above, the first term in (c) is the probability that the remainder of the run containing $X_{a_{j-1}}$ is completely deleted, the second term is the probability that the next k runs are deleted (k may be equal to zero), and the third term is the probability that the subsequent run is not completely deleted. This completes the proof of the lemma. \square

We now show that $\lim_{n \rightarrow \infty} \frac{1}{n} H_P(S^{M_n+1} | Y^{M_n})$ and $\lim_{n \rightarrow \infty} \frac{1}{n} H_P(X^n | Y^{M_n}, S^{M_n+1})$ each exist, thereby proving the existence of $\lim_{n \rightarrow \infty} \frac{1}{n} H_P(X^n, S^{M_n+1} | Y^{M_n})$.

Lemma 5.1. $\lim_{n \rightarrow \infty} \frac{1}{n} H_P(S^{M_n+1} | Y^{M_n}) = (1-d) H_P(S_2 | Y_1 Y_2)$ where the joint distribution of (Y_1, Y_2, S_2) is given by (28), (29), and (30).

Proof. See Appendix B.1. \square

To determine the limiting behavior of $\frac{1}{n} H(X^n | S^{M_n+1}, Y^{M_n})$, we recall that X^n can be equivalently represented in terms of its run-lengths as $(L_1^X, \dots, L_{R_n}^X)$, where R_n , the number of runs in X^n , is a random variable. Also recall from the discussion at the beginning of this section that the pair of sequences (S^{M_n+1}, Y^{M_n}) is equivalent to an augmented sequence \underline{Y}' formed by adding the positions of the deleted runs to $\underline{Y} = Y^{M_n}$. \underline{Y}' can be equivalently represented in terms of its run-lengths as $(L_1^{Y'}, \dots, L_{R_n}^{Y'})$, where we emphasize that $L_1^{Y'}, L_2^{Y'}, \dots$ can take value 0 as well. To summarize, we have

$$\begin{aligned}
X^n & \leftrightarrow (L_1^X, \dots, L_{R_n}^X), \\
(S^{M_n+1}, Y^{M_n}) & \leftrightarrow (L_1^{Y'}, \dots, L_{R_n}^{Y'}).
\end{aligned} \tag{34}$$

Thus, for all n

$$H_P(X^n | S^{M_n+1}, Y^{M_n}) = H_P(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'}). \tag{35}$$

Proposition 5.3. *The process $\{\mathbf{L}^{\mathbf{X}}, \mathbf{L}^{\mathbf{Y}'}\} \triangleq \{(L_1^X, L_1^{Y'}), (L_2^X, L_2^{Y'}), \dots\}$ is an i.i.d process characterized by the following joint distribution for all $j \geq 1$:*

$$P(L_j^X = r, L_j^{Y'} = s) = \gamma^{r-1}(1-\gamma) \cdot \binom{r}{s} d^{r-s}(1-d)^s, \quad r = 1, 2, \dots, 0 \leq s \leq r. \quad (36)$$

Proof. Since \mathbf{X} is a Markov process, $\{L_j^X\}_{j \geq 1}$ are independent with

$$P(L_j^X = r) = \gamma^{r-1}(1-\gamma), \quad r = 1, 2, \dots$$

Since the deletion process is i.i.d, each $L_j^{Y'}$ can be thought of being obtained by passing a run of length L_j^X through a discrete memoryless channel with transition probability

$$P(L_j^{Y'} = s | L_j^X = r) = \binom{r}{s} d^{r-s}(1-d)^s, \quad 0 \leq s \leq r.$$

□

Lemma 5.2. $\lim_{n \rightarrow \infty} \frac{1}{n} H_P(X^n | S^{M_n+1}, Y^{M_n}) = (1-\gamma) H_P(L^X | L^{Y'})$ where the joint distribution of $(L^X, L^{Y'})$ is given by (36).

Proof. See Appendix B.2. □

Using Lemmas 5.1 and 5.2, we obtain the following lower bound on the capacity of the deletion channel.

Theorem 3. *The deletion channel capacity $C(d)$ can be lower bounded as*

$$C(d) \geq \max_{0 < \gamma < 1} h(\gamma) - (1-d)H(S_2 | Y_1 Y_2) - (1-\gamma)H(L^X | L^{Y'})$$

where

$$H(S_2 | Y_1 Y_2) = \gamma \bar{\theta} \log_2 \frac{q}{\bar{\theta}} + \frac{\beta \theta}{\bar{\theta}^2} \log_2 \frac{1}{\bar{\theta}} + \frac{\beta \theta}{1-\theta^2} \log_2 \frac{q}{\beta} + \frac{\beta}{1-\theta^2} \log_2 \frac{\bar{q}}{\beta}, \quad (37)$$

$$q = \frac{\gamma + d - 2\gamma d}{1 + d - 2\gamma d}, \quad \theta = \frac{(1-\gamma)d}{1-\gamma d}, \quad \beta = \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2}$$

and

$$\begin{aligned} H(L^X | L^{Y'}) &= \left(\frac{d}{\bar{\gamma}} - \frac{d\bar{\gamma}}{(1-\gamma d)^2} \right) \log_2 \frac{1}{\gamma d} + \frac{d\bar{\gamma}h(d\gamma)}{(1-d\gamma)^2} - \frac{\bar{d}(2-\gamma-\gamma d) \log_2(1-\gamma d)}{\bar{\gamma}(1-\gamma d)} \\ &\quad - \frac{\bar{\gamma}}{\gamma} \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} (\bar{d}\gamma)^k (d\gamma)^j \binom{j+k}{k} \log_2 \binom{j+k}{k}. \end{aligned} \quad (38)$$

Proof. Combining (4) and (27), we obtain

$$C(d) > h(\gamma) - \limsup_{n \rightarrow \infty} \frac{H_P(X^n, S^{M_n+1} | Y^{M_n})}{n}. \quad (39)$$

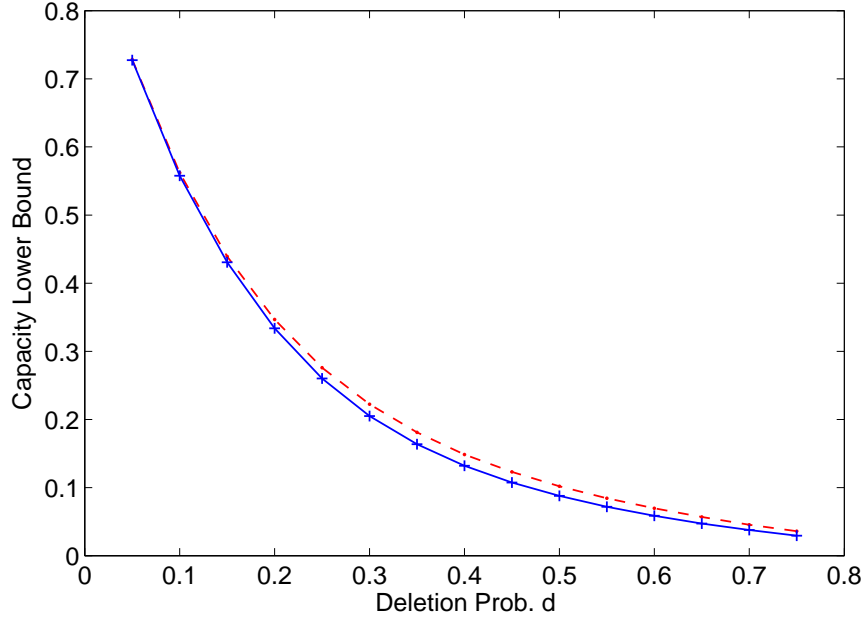


Figure 3: Lower bound of Theorem 3 on the deletion capacity. The lower bound from [5] is shown in dashed lines.

From Lemmas 5.1 and 5.2, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} H_P(X^n, S^{M_n+1} | Y^{M_n}) &= \lim_{n \rightarrow \infty} \frac{1}{n} H_P(S^{M_n+1} | Y^{M_n}) + \lim_{n \rightarrow \infty} \frac{1}{n} H_P(X^n | S^{M_n+1}, Y^{M_n}) \\ &= (1-d)H(S_2 | Y_1 Y_2) + (1-\gamma)H(L^X | L^{Y'}). \end{aligned} \quad (40)$$

$H(S_2 | Y_1 Y_2)$ can then be computed using the joint distribution given by (28), (29), and (30). $H(L^X | L^{Y'})$ can be computed using the joint distribution given in Proposition 5.3. Finally, we optimize the lower bound by maximizing over the Markov parameter $\gamma \in (0, 1)$. \square

Figure 3 shows the lower bound of Theorem 3 as well as the lower bound of [5] for various values of d . We observe that our bound is close to, but slightly smaller than that of [5], which is the best known lower bound on the deletion capacity.

In the decomposition of $H_P(X^n | Y^{M_n})$ in (26), we dropped the term $H_P(S^{M_n+1} | X^n, Y^{M_n})$ to obtain the bound in (27). $H_P(S^{M_n+1} | X^n, Y^{M_n})$ is the uncertainty in the positions of the deleted runs given both the input and output sequences. For example, if $X^n = 001100$ and $Y^{M_n} = 000$, then S^{M_n+1} equals

- $(0, 1, 0, 0)$ if the deletion pattern is either $0 \ 0 \ 1 \ 1 \ 0 \ 0$ or $0 \ 0 \ 1 \ 1 \ 0 \ 0$ (bits in italics are deleted).
- $(0, 0, 1, 0)$ if the deletion pattern is either $0 \ 0 \ 1 \ 1 \ 0 \ 0$ or $0 \ 0 \ 1 \ 1 \ 0 \ 0$.

Computing $\lim_{n \rightarrow \infty} H_P(S^{M_n+1} | X^n, Y^{M_n})$ precisely is hard, but obtaining a non-trivial lower bound for this quantity would improve the lower bound of Theorem 3.

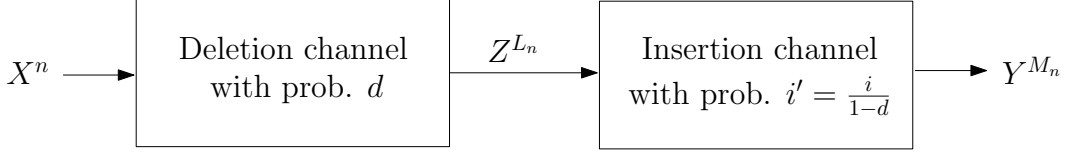


Figure 4: Cascade channel equivalent to the deletion+insertion channel.

6 The Deletion+Insertion Channel

Recall that this channel is defined by three parameters (d, i, α) . Each input bit undergoes a deletion with probability d , a duplication with probability $i\alpha$, a complementary insertion with probability $i\bar{\alpha}$. Note that each input bit is deleted with probability d ; given that a particular bit is *not* deleted, the probability that it undergoes an insertion is $\frac{i}{1-d}$. Therefore, one can think of the channel as a cascade of two channels, as shown in Figure 4. The first channel is a deletion channel that deletes each bit independently with probability d . The second channel is an insertion channel with parameters (i', α) , where $i' \triangleq \frac{i}{1-d}$. We prove the equivalence of this cascade decomposition below.

Claim: The deletion+insertion channel is thus equivalent to the cascade channel in the sense that both have the same transition probability $P(Y|X)$.

Proof. For an n -bit input sequence, define the deletion-insertion pattern $\Lambda^n = (\Lambda_1, \Lambda_2, \dots, \Lambda_n)$ of the channel as the sequence where Λ_i indicates whether the channel introduces a deletion/duplication/complementary insertion/no modification in bit i of the input. Note that if the underlying probability space is (Ω, \mathcal{F}, P) , the realization $\omega \in \Omega$ determines the deletion-insertion pattern $\Lambda^n(\omega)$. We calculate the probability of any specified pattern occurring in a) the deletion+insertion channel, and b) the cascade channel.

Consider a deletion-insertion pattern λ^n with k deletions at positions a_1, a_2, \dots, a_k , l duplications at positions b_1, \dots, b_l , and m complementary insertions at positions c_1, \dots, c_m . The probability of this pattern occurring in the deletion+insertion channel is

$$P_{delins}(\Lambda^n(\omega) = \lambda^n) = d^k (i\alpha)^l (i\bar{\alpha})^m (1 - d - i)^{n-k-l-m}.$$

The probability of this pattern occurring in the cascade channel of Figure 4 is

$$\begin{aligned} P_{casc}(\Lambda^n(\omega) = \lambda^n) &\stackrel{(a)}{=} [d^k (1-d)^{n-k}] \left[(i'\alpha)^l (i'\bar{\alpha})^m (1-i')^{(n-k)-l-m} \right] \\ &= [d^k (1-d)^{n-k}] \left[\left(\frac{i\alpha}{1-d} \right)^l \left(\frac{i\bar{\alpha}}{1-d} \right)^m \left(\frac{1-d-i}{1-d} \right)^{n-k-l-m} \right] \\ &= d^k (i\alpha)^l (i\bar{\alpha})^m (1-d-i)^{n-k-l-m}. \end{aligned} \tag{41}$$

where the first term in (a) is the probability of deletions occurring in the specified positions in the first channel, and the second term is the probability of the insertions occurring in the specified positions in the second channel. Hence every deletion-insertion pattern has the same probability in both the deletion+insertion channel and the cascade channel. This implies that the two channels have the same transition probability. \square

To obtain a lower bound on the capacity, we use two auxiliary sequences $T^{M_n} = (T_1, \dots, T_{M_n})$, and $S^{M_n+1} = (S_1, \dots, S_{M_n+1})$. As in Section 4.1, T^{M_n} indicates the complementary insertions in Y^{M_n} : $T_j = 1$ if

Y_j is a complementary insertion, and $T_j = 0$ otherwise. As in Section 5, S^{M_n+1} indicates the positions of the missing runs: $S_j = k$, if k runs were completely deleted between Y_{j-1} and Y_j . Using these auxiliary sequences, we can decompose $H_P(X^n|Y^{M_n})$ as

$$\begin{aligned} H_P(X^n|Y^{M_n}) &= H_P(X^n, T^{M_n}, S^{M_n+1}|Y^{M_n}) - H_P(T^{M_n}, S^{M_n+1}|X^n, Y^{M_n}) \\ &= H_P(T^{M_n}|Y^{M_n}) + H_P(S^{M_n+1}|T^{M_n}, Y^{M_n}) + H_P(X^n|S^{M_n+1}, T^{M_n}, Y^{M_n}) - H_P(T^{M_n}, S^{M_n+1}|X^n, Y^{M_n}) \\ &\leq H_P(T^{M_n}|Y^{M_n}) + H_P(S^{M_n+1}|T^{M_n}, Y^{M_n}) + H_P(X^n|S^{M_n+1}, \tilde{Y}^{M_n}) - H_P(T^{M_n}, S^{M_n+1}|X^n, Y^{M_n}) \end{aligned} \quad (42)$$

where \tilde{Y}^{M_n} is the sequence formed by flipping the complementary insertions in Y^{M_n} . The inequality in the last line of (42) holds because \tilde{Y}^{M_n} is a function of (T^{M_n}, Y^{M_n}) . We therefore have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n|Y^{M_n}) &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} (H_P(T^{M_n}|Y^{M_n}) + H_P(S^{M_n+1}|T^{M_n}, Y^{M_n}) + H_P(X^n|S^{M_n+1}, \tilde{Y}^{M_n})) \\ &\leq \limsup_{n \rightarrow \infty} \frac{H_P(T^{M_n}|Y^{M_n})}{n} + \limsup_{n \rightarrow \infty} \frac{H_P(S^{M_n+1}|T^{M_n}, Y^{M_n})}{n} + \limsup_{n \rightarrow \infty} \frac{H_P(X^n|S^{M_n+1}, \tilde{Y}^{M_n})}{n}. \end{aligned} \quad (43)$$

Using this upper bound for $\limsup_{n \rightarrow \infty} \frac{H_P(X^n|Y^{M_n})}{n}$ in (4), we obtain a lower bound on the capacity of the deletion+insertion channel.

Lemma 6.1. $\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(T^{M_n}|Y^{M_n}) = (1 - d + i) \limsup_{m \rightarrow \infty} \frac{1}{m} H_P(T^m|Y^m)$.

Proof. The proof follows the same steps as that of Lemma 4.1, with two changes: T^{M_n} replaces I^{M_n} , and we note that $\frac{M_n}{n}$ converges almost surely to $(1 - d + i)$ for the deletion+insertion channel. \square

Lemma 6.2. $\limsup_{m \rightarrow \infty} \frac{1}{m} H_P(T^m|Y^m) \leq \lim_{j \rightarrow \infty} H_P(T_j|T_{j-1}, Y_j, Y_{j-1})$, where

$$\lim_{j \rightarrow \infty} H_P(T_j|T_{j-1}, Y_j, Y_{j-1}) = \frac{(\bar{q}\bar{d} + qi\bar{\alpha})}{(1 - d + i)} h\left(\frac{i\bar{\alpha}}{\bar{q}\bar{d} + qi\bar{\alpha}}\right), \text{ and } q = \frac{\gamma + d - 2\gamma d}{1 + d - 2\gamma d}.$$

Proof. We have

$$\frac{H_P(T^m|Y^m)}{m} = \frac{\sum_{j=1}^m H_P(T_j|T_j^{j-1}, Y^m)}{m} \leq \frac{\sum_{j=1}^m H_P(T_j|T_{j-1}, Y_j, Y_{j-1})}{m}. \quad (44)$$

Therefore

$$\limsup_{m \rightarrow \infty} \frac{H_P(T^m|Y^m)}{m} \leq \limsup_{m \rightarrow \infty} \frac{\sum_{j=1}^m H_P(T_j|T_{j-1}, Y_j, Y_{j-1})}{m} = \lim_{j \rightarrow \infty} H_P(T_j|T_{j-1}, Y_j, Y_{j-1}), \quad (45)$$

provided the limit exists. From the cascade representation in Figure 4, we see that the insertions are introduced by the second channel in the cascade, an insertion channel with parameters (i', α) . The input to this insertion channel is a process $\mathbf{Z} = \{Z_m\}_{m \geq 1}$, which is the output of the first channel in the cascade. From Proposition 5.1, \mathbf{Z} is a first-order Markov process with parameter $q = \frac{\gamma + d - 2\gamma d}{1 + d - 2\gamma d}$.

Therefore, we need to calculate $\lim_{j \rightarrow \infty} H(T_j|T_{j-1}, Y_j, Y_{j-1})$ where \mathbf{Y} is the output when a first-order Markov process with parameter q is transmitted through an insertion channel with parameters (i', α) . But we have already computed $\lim_{j \rightarrow \infty} H(T_j|T_{j-1}, Y_j, Y_{j-1})$ in Lemma 4.6 for an insertion channel with parameters

(i, α) with a first-order Markov input with parameter γ . Hence, in Lemma 4.6, we can replace γ by q , and i by i' to obtain

$$\lim_{j \rightarrow \infty} H_P(T_j | T_{j-1}, Y_j, Y_{j-1}) = \frac{(1 - q + qi'\bar{\alpha})}{(1 + i')} h \left(\frac{i'\bar{\alpha}}{1 - q + qi'\bar{\alpha}} \right).$$

Substituting $i' = \frac{i}{1-d}$ and simplifying gives the statement of the lemma. \square

Lemma 6.3. $\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(S^{M_n+1} | T^{M_n}, Y^{M_n}) = (1 - d + i) \limsup_{m \rightarrow \infty} \frac{1}{m} H_P(S^m | T^m, Y^m).$

Proof. The proof is along the same lines as that of Lemma 4.1; here we use the uniform integrability of the sequence $\{-\frac{1}{n} \log P(S^{M_n+1} | T^{M_n}, Y^{M_n})\}$ along with the fact that $\frac{M_n}{n} \rightarrow (1 - d + i)$ almost surely. The uniform integrability follows from Lemma 2.2 since $\text{Supp}(S^{M_n+1} | T^{M_n}, Y^{M_n})$ is upper bounded by 2^n for the reasons explained in Section B.1. \square

Lemma 6.4. $\limsup_{m \rightarrow \infty} \frac{1}{m} H_P(S^m | T^m, Y^m) \leq \lim_{j \rightarrow \infty} H_P(S_j | Y_{j-1}, Y_j, T_j) = \frac{1}{1+i'} (A_1 + A_2 - \frac{\theta\beta}{(1-\theta)^2} \log_2 \theta),$ where

$$\begin{aligned} i' &= \frac{i}{1-d}, \quad q = \frac{\gamma + d - 2\gamma d}{1 + d - 2\gamma d}, \quad \theta = \frac{(1-\gamma)d}{1-\gamma d}, \quad \beta = \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2}, \\ A_1 &= \frac{\theta\beta(1-i'\bar{\alpha})}{1-\theta^2} \log_2 \left(\frac{i'\alpha + (1-i'\bar{\alpha})q + i'\bar{\alpha}\bar{q}}{\beta(1-i'\bar{\alpha})} \right) + \frac{\theta^2\beta i'\bar{\alpha}}{1-\theta^2} \log_2 \left(\frac{i'\alpha + (1-i'\bar{\alpha})q + i'\bar{\alpha}\bar{q}}{\beta i'\bar{\alpha}} \right) \\ &\quad + ((1-i'\bar{\alpha})\gamma\bar{\theta} + i'\bar{\alpha}\beta + i'\alpha) \log_2 \left(\frac{i'\alpha + (1-i'\bar{\alpha})q + i'\bar{\alpha}\bar{q}}{i'\alpha + (1-i'\bar{\alpha})\gamma\bar{\theta} + i'\bar{\alpha}\beta} \right), \\ A_2 &= \frac{\theta^2\beta(1-i'\bar{\alpha})}{1-\theta^2} \log_2 \left(\frac{(1-i'\bar{\alpha})\bar{q} + i'\bar{\alpha}q}{\beta(1-i'\bar{\alpha})} \right) + \frac{\theta\beta i'\bar{\alpha}}{1-\theta^2} \log_2 \left(\frac{(1-i'\bar{\alpha})\bar{q} + i'\bar{\alpha}q}{\beta i'\bar{\alpha}} \right) \\ &\quad + (i'\bar{\alpha}\gamma\bar{\theta} + (1-i'\bar{\alpha})\beta) \log_2 \left(\frac{(1-i'\bar{\alpha})\bar{q} + i'\bar{\alpha}q}{i'\bar{\alpha}\gamma\bar{\theta} + (1-i'\bar{\alpha})\beta} \right). \end{aligned}$$

Proof. See Appendix C.1 \square

We now determine the limiting behavior of $\frac{1}{n} H(X^n | S^{M_n+1}, \tilde{Y}^{M_n})$. By flipping the complementary insertions in Y^{M_n} to obtain \tilde{Y}^{M_n} , we have removed the extra runs introduced by the channel. Using S^{M_n+1} , we can augment \tilde{Y}^n by adding the positions of the deleted runs to obtain a sequence Y'^{M_n} which contains the same number of runs as X^n . Y'^{M_n} can be represented in terms of its run-lengths as $(L_1^{Y'}, \dots, L_{R_n}^{Y'})$, where we emphasize that $L_1^{Y'}, L_2^{Y'}, \dots$ can take value 0 as well. To summarize, we have

$$\begin{aligned} X^n &\leftrightarrow (L_1^X, \dots, L_{R_n}^X), \\ (S^{M_n+1}, \tilde{Y}^{M_n}) &\leftrightarrow (L_1^{Y'}, \dots, L_{R_n}^{Y'}). \end{aligned} \tag{46}$$

Thus, for all n

$$H_P(X^n | S^{M_n+1}, \tilde{Y}^{M_n}) = H_P(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'}). \tag{47}$$

Proposition 6.1. *The process $\{\mathbf{L}^X, \mathbf{L}^{Y'}\} \triangleq \{(L_1^X, L_1^{Y'}), (L_2^X, L_2^{Y'}), \dots\}$ is an i.i.d process characterized by the following joint distribution for all $j \geq 1$:*

$$\begin{aligned} P(L_j^X = r) &= \gamma^{r-1}(1-\gamma), \quad r = 1, 2, \dots \\ P(L_j^{Y'} = s | L_j^X = r) &= \sum_{n_i \in \mathcal{I}} \binom{r}{n_i; r + n_i - s} i^{n_i} d^{r+n_i-s} (1-d-i)^{s-2n_i}, \quad 0 \leq s \leq 2r \end{aligned} \tag{48}$$

where \mathcal{I} , the set of possible values for the number of insertions n_i , is given by

$$\mathcal{I} = \{0, 1, \dots, \lfloor \frac{s}{2} \rfloor\} \text{ for } s \leq r, \text{ and } \{s - r, \dots, \lfloor \frac{s}{2} \rfloor\} \text{ for } s > r.$$

Proof. Since \mathbf{X} is a Markov process, $\{L_j^X\}_{j \geq 1}$ are independent with

$$P(L_j^X = r) = \gamma^{r-1}(1 - \gamma), \quad r = 1, 2, \dots$$

Since there is a one-to-one correspondence between the runs of \mathbf{X} and the runs of \mathbf{Y}' , we can think of each $L_j^{Y'}$ being obtained by passing a run of length L_j^X through a discrete memoryless channel. For a pair $(L_j^X = r, L_j^{Y'} = s)$, if the number of insertions is n_i , the number of deletions is easily seen to be $r + n_i - s$. Since there can be at most one insertion after each input bit, no more than half the bits in an output run can be insertions; hence the maximum value of n_i is $\lfloor \frac{s}{2} \rfloor$. The minimum value of n_i is zero for $s \leq r$, and $s - r$ for $s > r$. Using these together with the fact that each bit can independently undergo an insertion with probability i , a deletion with probability d , or no change with probability $1 - d - i$, the transition probability of the memoryless run-length channel is given by the second line of (48). \square

Lemma 6.5. $\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(X^n | S^{M_n+1}, \tilde{Y}^{M_n}) = (1 - \gamma) H_P(L^X | L^{Y'})$ where the joint distribution of $(L^X, L^{Y'})$ is given by (48).

Proof. The proof is identical to that of Lemma 5.2. \square

Theorem 4. The capacity of the deletion+insertion channel can be lower bounded as

$$C(d, i, \alpha) \geq \max_{0 < \gamma < 1} h(\gamma) - (\bar{q}(1-d) + qi\bar{\alpha})h\left(\frac{i\bar{\alpha}}{\bar{q}(1-d) + qi\bar{\alpha}}\right) - (1-d)(A_1 + A_2 - \frac{\theta\beta}{(1-\theta)^2} \log_2 \theta) - (1-\gamma)H_P(L_1^X | L_1^{Y'})$$

where $q, \beta, \theta, A_1, A_2$ are defined in Lemma 6.4, and $H_P(L_1^X | L_1^{Y'})$ is computed using the joint distribution in (48).

Proof. The result is obtained by using Lemmas 6.1-6.5 in (43), and substituting the resulting bound in (4). \square

The lower bound is plotted in Figure 5 for various values of $d = i$, for $\alpha = 0.8$ and for $\alpha = 1$. For Theorem 4, we used the sequence T^{M_n} to indicate the positions of complementary insertions together with S^{M_n} to indicate deleted runs. We can obtain another lower bound on the deletion+insertion capacity by using the sequence I^{M_n} instead of T^{M_n} , in Section 4.1. This bound can be derived in a straightforward manner by combining the techniques of Sections 4.1 and 5, and is omitted.

7 Conclusion

The framework used in this paper suggests several directions for further progress on computing the capacity of channels with synchronization errors:

- There are a few different ways to sharpen the bounds for the insertion channel and the deletion+insertion channel. One target is the inequality in (19) - creating the sequence \tilde{Y}^{M_n} ensures one-to-one correspondence between input and output runs, but is not an optimal way to use the positions of complementary insertions. Is there a better way to use the knowledge of (T^{M_n}, Y^{M_n}) ?

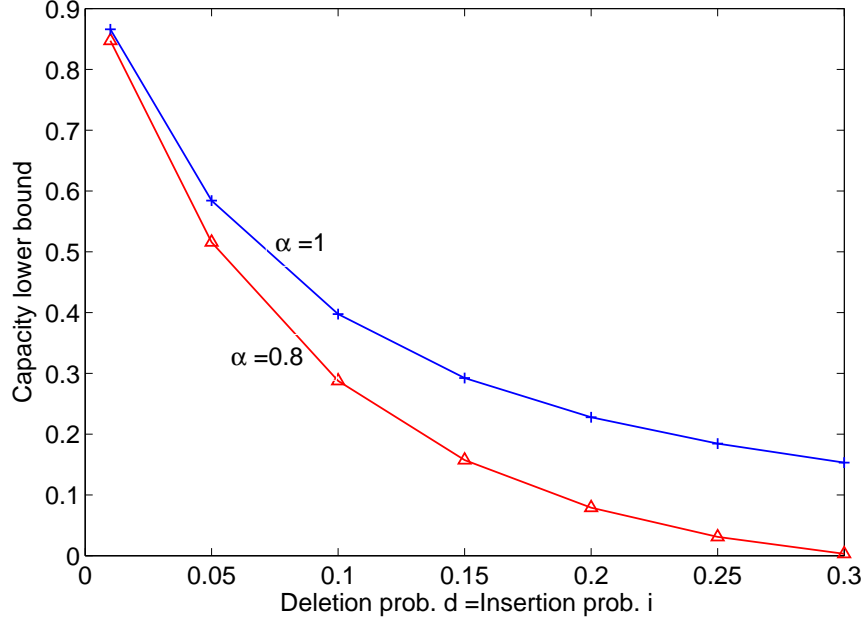


Figure 5: Lower bound on the deletion+insertion capacity $C(d, i, \alpha)$ for $d = i$.

When the input distribution is Markov, the deletion channel ensures that the output distribution is also Markov. But the presence of insertions results in an output process that is not Markov, which is the reason an exact expression for the limiting behavior of $\frac{1}{m}H(T^m|Y^m)$ could not be obtained in Lemma 4.6. A better bound for this term would improve the capacity lower bound.

- In the decomposition of $H(X^n|Y^{M_n})$ in (26) and (42), the penalty terms $-H(S^{M_n}|X^n, Y^{M_n})$ for the deletion channel, and $H(T^{M_n}, S^{M_n}|X^n, Y^{M_n})$ for the deletion+insertion channel - were dropped to obtain the capacity bounds. These terms are hard to compute precisely, but any non-trivial lower bound for these terms would improve the capacity bounds.
- Another direction is to investigate the performance of more general input distributions with i.i.d runs. For example, a distribution that is constant for small values and then decays geometrically may be a good run-length distribution for deletion channels since it decreases the probability of a run being completely deleted. A result on the structure of the optimal input distribution for small values of i and d (in the spirit of [10, 11]) would be very useful. Such a result could be combined with the approach used here to obtain good estimates of the capacity for small insertion and deletion probabilities.
- For the insertion channel, if we fix the insertion probability i and vary α , intuition suggests that the channel with $\alpha = 1$ has the largest capacity since there is always one-to-one correspondence between input and output runs, and no auxiliary sequences are needed. The capacity lower bound plotted in Figure 2 seems to verify this intuition. Formally proving this conjecture would yield an upper bound on the insertion capacity $C(i, \alpha)$ for $\alpha < 1$ since very tight bounds are known for the case of $\alpha = 1$ [12].
- One could also obtain upper bounds on the insertion capacity by considering a genie-aided decoder with access to the auxiliary sequence T^{M_n} , as done in [4] for the deletion channel. The task then is obtain an

upper bound on the capacity per unit cost of the equivalent DMC.

- The framework used here can be extended to derive bounds for channels with substitution errors in addition to deletions and insertions. For this, we would need an additional auxiliary sequence, e.g., a sequence that indicates the positions of the bit flips.
- The problem of synchronization also appears in file backup and file sharing [2,3], where distributed nodes with different versions of the same file want to synchronize their versions. For example, consider two nodes, with the first node having source \underline{X} and the second having source \underline{Y} , which is an edited version of \underline{X} . The edits may include deletions, insertions, and substitutions. A basic question is: To update \underline{Y} to \underline{X} , what is the minimum communication rate needed from the first node to the second? This is a distributed source coding problem, and it can be shown that the optimal rate is given by the limiting behavior of $H(\underline{X}|\underline{Y})$. The results derived in this paper provide bounds on this optimal rate for the case where \underline{X} is Markov, and the edit model $P(\underline{Y}|\underline{X})$ is one with i.i.d deletions and insertions. Extension of these results to edit models with substitution errors would yield rates to benchmark the performance of practical file synchronization tools such as rsync [18].

References

- [1] A. R. Iyengar, P. H. Siegel, and J. K. Wolf, “Write channel model for bit-patterned media recording,” *IEEE Transactions on Magnetics*, vol. 47, no. 1, pp. 35–45, 2011.
- [2] N. Ma, K. Ramchandran, and D. Tse, “Efficient file synchronization: A distributed source coding approach,” in *Proc. 2011 IEEE Int. Symp. Inf. Theory*, 2011.
- [3] R. Venkataramanan, S. Tatikonda, and K. Ramchandran, “Bounds on the optimal rate for synchronization from deletions and insertions,” in *Proc. Information Theory and Applications Workshop*, 2011.
- [4] S. N. Diggavi and M. Grossglauser, “On information transmission over a finite buffer channel,” *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1226–1237, 2006.
- [5] E. Drinea and M. Mitzenmacher, “Improved lower bounds for the capacity of i.i.d. deletion and duplication channels,” *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2693–2714, 2007.
- [6] A. Kirsch and E. Drinea, “Directly lower bounding the information capacity for channels with i.i.d. deletions and duplications,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 86–102, January 2010.
- [7] S. Diggavi, M. Mitzenmacher, and H. Pfister, “Capacity upper bounds for the deletion channel,” in *Proc. Int. Symp. on Inf. Theory*, 2007.
- [8] M. Mitzenmacher, “A survey of results for deletion channels and related synchronization channels,” *Probability Surveys*, vol. 6, pp. 1–33, 2009.
- [9] D. Fertionani and T. M. Duman, “Novel bounds on the capacity of the binary deletion channel,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 2753–2765, June 2010.
- [10] A. Kalai, M. Mitzenmacher, and M. Sudan, “Tight asymptotic bounds for the deletion channel with small deletion probabilities,” in *Proc. Int. Symp. on Inf. Theory*, 2010.

- [11] Y. Kanoria and A. Montanari, “On the deletion channel with small deletion probability,” in *Proc. Int. Symp. on Inf. Theory*, 2010.
- [12] M. Mitzenmacher, “Capacity bounds for sticky channels,” *IEEE Trans. on Inf. Theory*, vol. 54, pp. 72–77, Jan. 2008.
- [13] R. L. Dobrushin, “Shannon’s theorems for channels with synchronization errors,” *Problemy Peredachi Informatsii*, vol. 3, no. 4, pp. 18–36, 1967.
- [14] R. G. Gallager, “Sequential decoding for binary channels with noise and synchronization errors,” October 1961. Lincoln Lab Group Report.
- [15] D. Fertionani, T. M. Duman, and M. F. Erden, “Bounds on the capacity of channels with insertions, deletions and substitutions,” *IEEE Trans. on Communications*, vol. 59, no. 1, pp. 2–6, 2011.
- [16] P. Billingsley, *Probability and Measure*. Wiley-Interscience, 1995.
- [17] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 2006.
- [18] A. Tridgell and P. Mackerras, “The rsync algorithm.” <http://rsync.samba.org/>, Nov 1998.
- [19] P. H. Algoet and T. M. Cover, “A sandwich proof of the shannon-mcmillan-breiman theorem,” *Annals of Probability*, vol. 16, no. 2, pp. 899–909, 1988.

APPENDIX

A Insertion Channel

A.1 Proof of Lemma 4.1

We begin by noting that $\frac{M_n}{n} \rightarrow (1+i)$ almost surely, due to the strong law of large numbers. We have

$$\begin{aligned}
\frac{1}{n} H_P(I^{M_n} | Y^{M_n}) &= \mathbf{E} \left[-\frac{1}{n} \log P(I^{M_n} | Y^{M_n}) \cdot \left(\mathbf{1}_{\{\frac{M_n}{n} \in (1+i-\epsilon, 1+i+\epsilon)\}} + \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right) \right] \\
&= \mathbf{E} \left[-\frac{1}{n} \log \frac{P(I^{M_n}, Y^{M_n})}{P(Y^{M_n})} \cdot \mathbf{1}_{\{\frac{M_n}{n} \in (1+i-\epsilon, 1+i+\epsilon)\}} \right] + \mathbf{E} \left[-\frac{1}{n} \log P(I^{M_n} | Y^{M_n}) \cdot \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right] \\
&\leq \mathbf{E} \left[-\frac{1}{n} \log \frac{P(I^{n(1+i+\epsilon)}, Y^{n(1+i+\epsilon)})}{P(Y^{n(1+i+\epsilon)})} \cdot \mathbf{1}_{\{\frac{M_n}{n} \in (1+i-\epsilon, 1+i+\epsilon)\}} \right] + \mathbf{E} \left[-\frac{1}{n} \log P(I^{M_n} | Y^{M_n}) \cdot \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right] \\
&= \mathbf{E} \left[-\frac{1}{n} \log \frac{P(I^{n(1+i+\epsilon)}, Y^{n(1+i+\epsilon)})}{P(Y^{n(1+i+\epsilon)})} \right] - \mathbf{E} \left[-\frac{1}{n} \log \frac{P(I^{n(1+i+\epsilon)}, Y^{n(1+i+\epsilon)})}{P(Y^{n(1+i+\epsilon)})} \cdot \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right] \\
&\quad + \mathbf{E} \left[-\frac{1}{n} \log P(I^{M_n} | Y^{M_n}) \cdot \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right].
\end{aligned} \tag{49}$$

We first examine the third term in (49). The size of the support of $-\frac{1}{n} \log P(I^{M_n}|Y^{M_n})$ is at most 2^{2n} , since I^{M_n} is a binary sequence of length at most $2n$. Hence, from Lemma 2.2, $\{-\frac{1}{n} \log P(I^{M_n}|Y^{M_n})\}_{n \geq 1}$ is uniformly integrable. From Lemma 2.1, for any $\epsilon > 0$, there exists some $\delta > 0$

$$\mathbf{E} \left[-\frac{1}{n} \log P(I^{M_n}|Y^{M_n}) \cdot \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right] < \epsilon \quad (50)$$

whenever $\Pr(\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}) < \delta$. Since $\frac{M_n}{n} \rightarrow (1+i)$ almost surely, $\Pr(\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\})$ is less than δ for all sufficiently large n . Thus (50) is true for all sufficiently large n . Similarly, the third term can be shown to be smaller than ϵ for all sufficiently large n . Therefore, for all sufficiently large n , (49) becomes

$$\begin{aligned} \frac{1}{n} H_P(I^{M_n}|Y^{M_n}) &\leq \frac{H_P(I^{n(1+i+\epsilon)}, Y^{n(1+i+\epsilon)}) - H(Y^{n(1+i-\epsilon)})}{n} + \epsilon \\ &= (1+i-\epsilon) \frac{H_P(I^{n(1+i-\epsilon)}|Y^{n(1+i-\epsilon)})}{n(1+i-\epsilon)} + \frac{1}{n} H_P(I_{n(1+i-\epsilon)+1}^{n(1+i+\epsilon)}, Y_{n(1+i-\epsilon)+1}^{n(1+i+\epsilon)}|I^{n(1+i-\epsilon)}, Y^{n(1+i-\epsilon)}) + \epsilon \\ &\stackrel{(a)}{\leq} (1+i-\epsilon) \frac{H_P(I^{n(1+i-\epsilon)}|Y^{n(1+i-\epsilon)})}{n(1+i-\epsilon)} + 4\epsilon + \epsilon. \end{aligned} \quad (51)$$

where (a) holds because $I_{n(1+i-\epsilon)+1}^{n(1+i+\epsilon)}$ and $Y_{n(1+i-\epsilon)+1}^{n(1+i+\epsilon)}$ can each take on at most $2^{2n\epsilon}$ different values. Hence

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(I^{M_n}|Y^{M_n}) \leq 5\epsilon + (1+i+\epsilon) \limsup_{m \rightarrow \infty} \frac{1}{m} H_P(I^m|Y^m).$$

Since $\epsilon > 0$ is arbitrary, we let $\epsilon \rightarrow 0$ to obtain

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(I^{M_n}|Y^{M_n}) \leq (1+i) \limsup_{m \rightarrow \infty} \frac{1}{m} H_P(I^m|Y^m). \quad (52)$$

Using steps similar to (49), we have

$$\begin{aligned} \frac{1}{n} H_P(I^{M_n}|Y^{M_n}) &= \mathbf{E} \left[-\frac{1}{n} \log \frac{P(I^{M_n}, Y^{M_n})}{P(Y^{M_n})} \cdot \mathbf{1}_{\{\frac{M_n}{n} \in (1+i-\epsilon, 1+i+\epsilon)\}} \right] + \mathbf{E} \left[-\frac{1}{n} \log P(I^{M_n}|Y^{M_n}) \cdot \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right] \\ &\geq \mathbf{E} \left[-\frac{1}{n} \log \frac{P(I^{n(1+i-\epsilon)}, Y^{n(1+i-\epsilon)})}{P(Y^{n(1+i-\epsilon)})} \cdot \mathbf{1}_{\{\frac{M_n}{n} \in (1+i-\epsilon, 1+i+\epsilon)\}} \right] + \mathbf{E} \left[-\frac{1}{n} \log P(I^{M_n}|Y^{M_n}) \cdot \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right] \\ &= \mathbf{E} \left[-\frac{1}{n} \log \frac{P(I^{n(1+i-\epsilon)}, Y^{n(1+i-\epsilon)})}{P(Y^{n(1+i-\epsilon)})} \right] - \mathbf{E} \left[-\frac{1}{n} \log \frac{P(I^{n(1+i-\epsilon)}, Y^{n(1+i-\epsilon)})}{P(Y^{n(1+i-\epsilon)})} \cdot \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right] \\ &\quad + \mathbf{E} \left[-\frac{1}{n} \log P(I^{M_n}|Y^{M_n}) \cdot \mathbf{1}_{\{\frac{M_n}{n} \notin (1+i-\epsilon, 1+i+\epsilon)\}} \right]. \end{aligned} \quad (53)$$

Using arguments identical to the ones following (49), one can show that the last two terms of (53) are smaller than ϵ in absolute value for all sufficiently large n , leading to

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H_P(I^{M_n}|Y^{M_n}) \geq (1+i) \limsup_{m \rightarrow \infty} \frac{1}{m} H_P(I^m|Y^m). \quad (54)$$

Combining (52) and (54) completes the proof of the lemma. \square

A.2 Proof of Lemma 4.2

We have

$$\frac{1}{m} H_P(I^m | Y^m) = \frac{1}{m} \sum_{j=1}^m H_P(I_j | I^{j-1}, Y^m) \leq \frac{1}{m} \sum_{j=1}^m H_P(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2}) \quad (55)$$

where the inequality holds because conditioning cannot increase the entropy. Therefore

$$\limsup_{m \rightarrow \infty} \frac{1}{m} H_P(I^m | Y^m) \leq \limsup_{m \rightarrow \infty} \frac{1}{m} \sum_{j=1}^m H_P(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2}) = \lim_{j \rightarrow \infty} H_P(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2}), \quad (56)$$

provided the limit exists. We now show that $\lim_{j \rightarrow \infty} H_P(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2})$ exists and is given by (18).

From Proposition 4.1, the process $\{\mathbf{I}, \mathbf{Y}\}$ is characterized by a Markov chain with state at time j given by (I_j, Y_j, Y_{j-1}) . For any $\epsilon > 0$, the distribution $P(I_j, Y_j, Y_{j-1})$ is at most ϵ (in total variation norm) from the stationary joint distribution π given by (17) for all sufficiently large j . The conditional distribution $P(I_j, Y_j | I_{j-1}, Y_{j-2}, Y_{j-1})$ is given by (16). Due to the continuity of the entropy function, this implies

$$\lim_{j \rightarrow \infty} H_P(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2}) = \lim_{j \rightarrow \infty} H_\pi(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2}).$$

where π refers to the stationary joint distribution on $(Y_{j-2}, Y_{j-1}, I_{j-1}, I_j, Y_j)$, given by (17) and (16).

$H_\pi(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2})$ can be computed as follows. First, we note that $I_j = 0$ whenever $I_{j-1} = 1$. Therefore

$$\begin{aligned} H_\pi(I_j | I_{j-1}, Y_j, Y_{j-1}, Y_{j-2}) &= \sum_{y=0}^1 \pi(I_{j-1} = 0, Y_j = y, Y_{j-1} = y, Y_{j-2} = y) H(I_j | I_{j-1} = 0, Y_j = y, Y_{j-1} = y, Y_{j-2} = y) \\ &\quad + \pi(I_{j-1} = 0, Y_j = y, Y_{j-1} = y, Y_{j-2} = \bar{y}) H(I_j | I_{j-1} = 0, Y_j = y, Y_{j-1} = y, Y_{j-2} = \bar{y}) \\ &\quad + \pi(I_{j-1} = 0, Y_j = y, Y_{j-1} = \bar{y}, Y_{j-2} = \bar{y}) H(I_j | I_{j-1} = 0, Y_j = y, Y_{j-1} = \bar{y}, Y_{j-2} = \bar{y}) \\ &\quad + \pi(I_{j-1} = 0, Y_j = y, Y_{j-1} = \bar{y}, Y_{j-2} = y) H(I_j | I_{j-1} = 0, Y_j = y, Y_{j-1} = \bar{y}, Y_{j-2} = y) \end{aligned} \quad (57)$$

From (17) and (16), we have

$$\begin{aligned} \pi(I_{j-1} = 0, Y_j = y, Y_{j-1} = y, Y_{j-2} = y) &= \pi(I_{j-1} = 0, Y_{j-1} = y, Y_{j-2} = y) [P(Y_j = y, I_j = 1 | I_{j-1} = 0, Y_{j-1} = y, Y_{j-2} = y) \\ &\quad + P(Y_j = y, I_j = 0 | I_{j-1} = 0, Y_{j-1} = y, Y_{j-2} = y)] \\ &= \frac{\bar{i}\gamma + i\alpha\gamma + i\bar{\alpha}\bar{\gamma}}{2(1+i)} \cdot (i\alpha + \bar{i}\gamma), \end{aligned} \quad (58)$$

and

$$H(I_j | I_{j-1} = 0, Y_j = y, Y_{j-1} = y, Y_{j-2} = y) = h\left(\frac{i\alpha}{i\alpha + \bar{i}\gamma}\right). \quad (59)$$

The remaining terms in (57) can be similarly calculated to obtain (18). \square

A.3 Proof of Lemma 4.4

We have

$$\begin{aligned}
H_P(I^{n(1+i)}|Y^{n(1+i)}, X^n) &= \sum_{j=2}^{n(1+i)} H_P(I_{j+1}|I^j, Y^{n(1+i)}, X^n) \\
&\stackrel{(a)}{\geq} \sum_{j=1}^{n(1+i)} \sum_{y \in \{0,1\}} P(I^{j-1}, I_j = 0) P(Y^{j-1}, (Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y) | I_j = 0, I^{j-1}) \\
&\quad \cdot H_P(I_{j+1}|Y^{j-1}, I^{j-1}, I_j = 0, (Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y))
\end{aligned} \tag{60}$$

where k_j is the index of the input bit that corresponds to Y_j . (k_j is uniquely determined given I_1, \dots, I_j .) (a) is obtained as follows. Since $I_{j+1} = 0$ whenever $I_j = 1$, the entropy terms in the sum are non-zero only when $I_j = 0$. The inequality appears because we sum only over those indices j that satisfy

$$I_j = 0, (Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y).$$

For such indices, the input bit X_{k_j} corresponds to Y_j , and X_{k_j+2} to Y_{j+3} , the uncertainty in I_{j+1} being whether X_{k_j+1} corresponds to Y_{j+1} or Y_{j+2} . We have

$$\begin{aligned}
&P((Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y) | I_j = 0) \\
&= P((X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y)) \cdot P((Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y) | (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y), I_j = 0) \\
&= \frac{1}{2} \bar{\gamma}^2 \cdot (i\bar{\alpha} + (1-i)i\alpha)
\end{aligned} \tag{61}$$

where term $i\bar{\alpha}$ corresponds to the case where Y_{j+1} is a complementary insertion ($I_{j+1} = 1, I_{j+2} = 0$), and the term $(1-i)i\alpha$ to the case where Y_{j+2} is a duplication ($I_{j+1} = 0, I_{j+2} = 1$). Therefore,

$$H_P(I_{j+1}|I^{j-1}, I_j = 0, (Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y)) = h\left(\frac{i\bar{\alpha}}{i\bar{\alpha} + (1-i)i\alpha}\right). \tag{62}$$

As explained in Section 4, $\{I_j\}_{j \geq 1}$, is a Markov chain with $P(I_j = 0)$ converging to $\frac{1}{1+i}$ as $j \rightarrow \infty$. Substituting this along with (61) and (62) in (60), we obtain

$$\liminf_{n \rightarrow \infty} \frac{1}{n} H_P(I^{n(1+i)}|Y^{n(1+i)}, X^n) \geq n(1+i) \cdot \frac{1}{1+i} \bar{\gamma}^2 \cdot (i\bar{\alpha} + (1-i)i\alpha) h\left(\frac{\bar{\alpha}}{\bar{\alpha} + (1-i)\alpha}\right).$$

□

A.4 Proof of Lemma 4.6

We have

$$\frac{1}{m} H_P(T^m|Y^m) = \frac{1}{m} \sum_{j=1}^m H_P(T_j|T^{j-1}, Y^m) \leq \frac{1}{m} \sum_{j=1}^m H_P(T_j|T_{j-1}, Y_j, Y_{j-1}). \tag{63}$$

Therefore

$$\limsup_{m \rightarrow \infty} \frac{1}{m} H_P(T^m | Y^m) \leq \limsup_{m \rightarrow \infty} \frac{1}{m} \sum_{j=1}^m H_P(T_j | T_{j-1}, Y_j, Y_{j-1}) = \lim_{j \rightarrow \infty} H_P(T_j | T_{j-1}, Y_j, Y_{j-1}), \quad (64)$$

provided the limit exists. We now show that $\lim_{j \rightarrow \infty} H_P(T_j | T_{j-1}, Y_j, Y_{j-1})$ exists and is given by (22).

Note that $T_j = 0$ whenever $T_{j-1} = 1$ since we cannot have two consecutive insertions. Also, $T_j = 0$ whenever $Y_j = Y_{j-1}$ since $T_j = 1$ only when Y_j is a complementary insertion. Thus we have for all $j \geq 2$:

$$\begin{aligned} H(T_j | T_{j-1}, Y_j, Y_{j-1}) = & P(T_{j-1} = 0, Y_j = 1, Y_{j-1} = 0) H(T_j | T_{j-1} = 0, Y_j = 1, Y_{j-1} = 0) \\ & + P(T_{j-1} = 0, Y_j = 0, Y_{j-1} = 1) H(T_j | T_{j-1} = 0, Y_j = 0, Y_{j-1} = 1). \end{aligned} \quad (65)$$

Note that for all $j \geq 1$, $P(T_j = 1) = P(I_j = 1)\bar{\alpha}$, where $I_j = 1$ if Y_j is an inserted bit, and $I_j = 0$ otherwise. Therefore,

$$P(T_j = 0) = 1 - P(I_j = 1)\bar{\alpha}, \quad j \geq 1. \quad (66)$$

Note that the binary-valued process $\{I_j\}_{j \geq 1}$ is a Markov chain with transition probabilities

$$\Pr(I_j = 1 | I_j = 0) = 1 - \Pr(I_j = 0 | I_j = 0) = i, \quad \Pr(I_j = 1 | I_j = 1) = 1 - \Pr(I_j = 0 | I_j = 1) = 0. \quad (67)$$

For $i \in (0, 1)$, this is an irreducible, aperiodic Markov chain. Hence a unique stationary distribution π exists, which is given by

$$\pi(I_j = 1) = 1 - \pi(I_j = 0) = \frac{i}{1+i}. \quad (68)$$

Hence for any $\epsilon > 0$,

$$\left| P(I_j = 1) - \frac{i}{1+i} \right| < \epsilon \text{ and } \left| \Pr(I_j = 0) - \frac{1}{1+i} \right| < \epsilon \quad (69)$$

for all sufficiently large j . Using this in (66), for all sufficiently large j , the distribution $P(T_j)$ is within total variation norm ϵ of the following stationary distribution.

$$\pi(T_j = 0) = 1 - \frac{i\bar{\alpha}}{1+i} = \frac{1+i\alpha}{1+i}, \quad \pi(T_j = 1) = \frac{i\bar{\alpha}}{1+i}. \quad (70)$$

Further, we have $P(Y_j = 1 | T_j = 0) = P(Y_j = \bar{y} | T_j = 0) = 0.5$, for $y \in \{0, 1\}$ since both the input distribution and the insertion process are symmetric in 0 and 1. Hence the stationary distribution for (T_{j-1}, Y_{j-1}) is

$$\pi(T_{j-1} = 0, Y_{j-1} = y) = \frac{1+i\alpha}{2(1+i)}, \quad \pi(T_{j-1} = 1, Y_{j-1} = y) = \frac{i\bar{\alpha}}{2(1+i)}. \quad (71)$$

Next, we determine the conditional distribution $P(Y_j, T_j | Y_{j-1} = y, T_{j-1} = 0)$ for $y \in \{0, 1\}$. We have

$$\begin{aligned}
& P(T_j = 0, Y_j = y | Y_{j-1} = y, T_{j-1} = 0) \\
&= P(T_j = 0, Y_j = y, I_{j-1} = 1 | Y_{j-1} = y, T_{j-1} = 0) + P(T_j = 0, Y_j = y, I_{j-1} = 0 | Y_{j-1} = y, T_{j-1} = 0) \\
&= P(I_{j-1} = 1 | T_{j-1} = 0) \cdot P(T_j = 0, Y_j = y | I_{j-1} = 1, T_{j-1} = 0, Y_{j-1} = y) \\
&\quad + P(I_{j-1} = 0 | T_{j-1} = 0) \cdot P(T_j = 0, Y_j = y | I_{j-1} = 0, T_{j-1} = 0, Y_{j-1} = y) \\
&\stackrel{(a)}{=} \frac{P(I_{j-1} = 1)P(T_{j-1} = 0 | I_{j-1} = 1)}{P(T_{j-1} = 0)}\gamma + \frac{P(I_{j-1} = 0)P(T_{j-1} = 0 | I_{j-1} = 0)}{P(T_{j-1} = 0)}((1-i)\gamma + i\alpha) \\
&\stackrel{(b)}{=} \frac{P(I_{j-1} = 1)\alpha}{1 - \bar{\alpha}P(I_{j-1} = 1)}\gamma + \frac{P(I_{j-1} = 0)}{1 - \bar{\alpha}P(I_{j-1} = 1)}((1-i)\gamma + i\alpha).
\end{aligned} \tag{72}$$

In the above, (b) is obtained using (66). (a) is obtained as follows. The event $(I_{j-1} = 1, T_{j-1} = 0, Y_{j-1} = y)$ implies Y_{j-1} is a duplication, and hence $Y_{j-2} = y$ corresponds to an input bit (say X_a), and Y_j is the next input bit X_{a+1} . The probability that $X_{a+1} = X_a$ is γ . Hence $P(T_j = 0, Y_j = y | I_{j-1} = 1, T_{j-1} = 0, Y_{j-1} = y) = \gamma$. When $(I_{j-1} = 0, T_{j-1} = 0, Y_{j-1} = y)$, Y_{j-1} corresponds to an input bit, say X_b . Conditioned on this, the event $(T_j = 0, Y_j = y)$ can occur in two ways:

- Y_j is the next input bit X_{b+1} and is equal to y . This event has probability $(1-i)\gamma$.
- Y_j is a duplication of Y_{j-1} . This event has probability $i\alpha$.

Hence $P(T_j = 0, Y_j = y | I_{j-1} = 0, T_{j-1} = 0, Y_{j-1} = y) = ((1-i)\gamma + i\alpha)$. Similarly, we calculate

$$\begin{aligned}
& P(T_j = 0, Y_j = \bar{y} | Y_{j-1} = y, T_{j-1} = 0) = P(I_{j-1} = 1 | T_{j-1} = 0)P(T_j = 0, Y_j = \bar{y} | I_{j-1} = 1, T_{j-1} = 0, Y_{j-1} = y) \\
&\quad + P(I_{j-1} = 0 | T_{j-1} = 0)P(T_j = 0, Y_j = \bar{y} | I_{j-1} = 0, T_{j-1} = 0, Y_{j-1} = y) \\
&= \frac{P(I_{j-1} = 1)P(T_{j-1} = 0 | I_{j-1} = 1)}{P(T_{j-1} = 0)}\bar{\gamma} + \frac{P(I_{j-1} = 0)P(T_{j-1} = 0 | I_{j-1} = 0)}{P(T_{j-1} = 0)}(1-i)\bar{\gamma} \\
&= \frac{P(I_{j-1} = 1)}{1 - \bar{\alpha}P(I_{j-1} = 1)}\alpha\bar{\gamma} + \frac{P(I_{j-1} = 0)}{1 - \bar{\alpha}P(I_{j-1} = 1)}(1-i)\bar{\gamma},
\end{aligned} \tag{73}$$

$$\begin{aligned}
& P(T_j = 1, Y_j = \bar{y} | Y_{j-1} = y, T_{j-1} = 0) = P(I_{j-1} = 1 | T_{j-1} = 0)P(T_j = 1, Y_j = \bar{y} | I_{j-1} = 1, T_{j-1} = 0, Y_{j-1} = y) \\
&\quad + P(I_{j-1} = 0 | T_{j-1} = 0)P(T_j = 1, Y_j = \bar{y} | I_{j-1} = 0, T_{j-1} = 0, Y_{j-1} = y) \\
&= \frac{P(I_{j-1} = 1)P(T_{j-1} = 0 | I_{j-1} = 1)}{P(T_{j-1} = 0)}0 + \frac{P(I_{j-1} = 0)P(T_{j-1} = 0 | I_{j-1} = 0)}{P(T_{j-1} = 0)}i\bar{\alpha} \\
&= \frac{P(I_{j-1} = 0)}{1 - \bar{\alpha}P(I_{j-1} = 1)}i\bar{\alpha},
\end{aligned} \tag{74}$$

and

$$P(T_j = 1, Y_j = y | Y_{j-1} = y, T_{j-1} = 0) = 0. \tag{75}$$

Using (69) in equations (72)-(75), we see that for all sufficiently large j , the distribution $P(T_j, Y_j | Y_{j-1} =$

$y, T_{j-1} = 0$) is within a total variation norm ϵ from the following stationary distribution

$$\begin{aligned}
\pi(T_j = 0, Y_j = y | Y_{j-1} = y, T_{j-1} = 0) &= \frac{i\alpha(1 + \gamma) + (1 - i)\gamma}{1 + i\alpha}, \\
\pi(T_j = 0, Y_j = \bar{y} | Y_{j-1} = y, T_{j-1} = 0) &= \frac{\bar{\gamma}(1 - i\bar{\alpha})}{1 + i\alpha}, \\
\pi(T_j = 1, Y_j = \bar{y} | Y_{j-1} = y, T_{j-1} = 0) &= \frac{i\bar{\alpha}}{1 + i\alpha}, \\
\pi(T_j = 1, Y_j = y | Y_{j-1} = y, T_{j-1} = 0) &= 0.
\end{aligned} \tag{76}$$

Due to the continuity of the entropy function in the joint distribution, we therefore have

$$\lim_{j \rightarrow \infty} H_P(T_j | T_{j-1}, Y_{j-1}, Y_j) = H_\pi(T_j | T_{j-1}, Y_{j-1}, Y_j),$$

where the joint distribution $\pi(T_{j-1}, Y_{j-1}, Y_j, T_j)$ is given by (71) and (76). Using this in (65), one can compute $H_\pi(T_j | T_{j-1}, Y_{j-1}, Y_j)$ to obtain the result in the lemma. \square

A.5 Proof of Lemma 4.7

Due to (23), it is enough to show that $\frac{1}{n} H_P(L_1^X, \dots, L_{R_n}^X | L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}})$ converges to $(1 - \gamma) H_P(L_1^X | L_1^{\tilde{Y}})$. Since $\{(L_1^X, L_1^{\tilde{Y}}), (L_2^X, L_2^{\tilde{Y}}), \dots\}$ is an i.i.d process, from the strong law of large numbers, we have

$$\lim_{m \rightarrow \infty} \frac{1}{m} \log \Pr(L_1^X, \dots, L_m^X | L_1^{\tilde{Y}}, \dots, L_m^{\tilde{Y}}) = H_P(L_1^X | L_1^{\tilde{Y}}) \quad a.s. \tag{77}$$

Further, we have the normalized number of input runs $\frac{R_n}{n} \rightarrow (1 - \gamma)$ almost surely. Using the above in Lemma 2.4, we obtain

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}}) = H_P(L_1^X | L_1^{\tilde{Y}}) \quad a.s. \tag{78}$$

We now argue that $-\frac{1}{n} \log \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}})$ is uniformly integrable. $\text{Supp}(L_1^X, \dots, L_{R_n}^X | L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}})$ can be upper bounded by 2^n since the random sequence $(L_1^X, \dots, L_{R_n}^X)$ is equivalent to X^n , which can take on at most 2^n values. Hence, from Lemma 2.2, $-\frac{1}{n} \log \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}})$ is uniformly integrable. Using this together with (78) in Lemma 2.3, we conclude that

$$\begin{aligned}
\lim_{n \rightarrow \infty} \frac{1}{n} H_P(L_1^X, \dots, L_{R_n}^X | L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}}) &= \lim_{n \rightarrow \infty} \mathbf{E} \left[-\frac{1}{n} \log \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}}) \right] \\
&= \mathbf{E} \left[\lim_{n \rightarrow \infty} -\frac{1}{n} \log \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{\tilde{Y}}, \dots, L_{R_n}^{\tilde{Y}}) \right] \\
&= H_P(L_1^X | L_1^{\tilde{Y}}).
\end{aligned} \tag{79}$$

\square

A.6 Proof of Lemma 4.9

We have

$$\begin{aligned}
H_P(T^{n(1+i)}|Y^{n(1+i)}, X^n) &= \sum_{j=2}^{n(1+i)} H_P(T_{j+1}|T^j, Y^{n(1+i)}, X^n) \\
&\geq \sum_{j=2}^{n(1+i)} H_P(T_{j+1}|I_j, T^j, Y^{n(1+i)}, X^n) \\
&\stackrel{(a)}{\geq} \sum_{j=1}^{n(1+i)} \sum_{y \in \{0,1\}} P(T^{j-1}, I_j = 0) P(Y^{j-1}, (Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y) | I_j = 0, T^{j-1}) \\
&\quad \cdot H_P(T_{j+1}|Y^{j-1}, T^{j-1}, I_j = 0, (Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y)).
\end{aligned} \tag{80}$$

(a) is obtained as follows. Since $T_{j+1} = 0$ whenever $I_j = 1$, the entropy terms in the second line are non-zero only when $I_j = 0$. The inequality appears because we only sum over indices j such that

$$I_j = 0, (Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y),$$

where k_j is the index of the input bit that corresponds to Y_j . This is uniquely determined because given T^j , there is a one-to-one correspondence between the input and output runs until bit Y^j . Then $Y_j = y$ and $Y_{j+1} = \bar{y}$ implies k_j is the last input bit in the run containing Y_j . Therefore, input bit X_{k_j} corresponds to Y_j , and X_{k_j+2} corresponds to Y_{j+3} , the uncertainty being whether X_{k_j+1} corresponds to Y_{j+1} or Y_{j+2} . We have

$$\begin{aligned}
&P((Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y) | I_j = 0) \\
&= P((X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y) | I_j = 0) \cdot P((Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y) | (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y), I_j = 0) \\
&= \frac{1}{2} \bar{\gamma}^2 \cdot (i\bar{\alpha} + (1-i)i\alpha)
\end{aligned} \tag{81}$$

where term $i\bar{\alpha}$ corresponds to the case where Y_{j+1} is a complementary insertion ($T_{j+1} = 1$), and the term $(1-i)i\alpha$ to the case where Y_{j+2} is a duplication ($T_{j+1} = 0$). Consequently,

$$H_P(T_{j+1}|T^{j-1}, I_j = 0, (Y_j, Y_{j+1}, Y_{j+2}, Y_{j+3}) = (y, \bar{y}, \bar{y}, y), (X_{k_j}, X_{k_j+1}, X_{k_j+2}) = (y, \bar{y}, y)) = h\left(\frac{i\bar{\alpha}}{i\bar{\alpha} + (1-i)i\alpha}\right). \tag{82}$$

Substituting (81) and (82) in (80) and using the fact that $P(I_j = 0) \rightarrow \frac{1}{1+i}$, we obtain

$$\liminf_{n \rightarrow \infty} \frac{1}{n} H_P(T^{n(1+i)}|Y^{n(1+i)}, X^n) \geq n(1+i) \cdot \frac{1}{1+i} \bar{\gamma}^2 \cdot (i\bar{\alpha} + (1-i)i\alpha) h\left(\frac{\bar{\alpha}}{\bar{\alpha} + (1-i)\alpha}\right).$$

□

B Deletion Channel

B.1 Proof of Lemma 5.1

We first show that almost surely

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log P(S^{M_n} | Y^{M_n}) = (1-d)H_P(S_2 | Y_1 Y_2).$$

From Propositions 5.1 and 5.2, $\{Y_m\}_{m \geq 1}$ and $\{(S_m, Y_m)\}_{m \geq 1}$ are both ergodic Markov chains with stationary transition probabilities. Therefore, from the Shannon-McMillan-Breiman theorem [19], we have

$$\lim_{m \rightarrow \infty} -\frac{1}{m} \log P(Y^m) = H_P(Y_2 | Y_1) \text{ a.s.}, \quad (83)$$

$$\lim_{m \rightarrow \infty} -\frac{1}{m} \log P(S^m, Y^m) = H_P(S_2, Y_2 | Y_1) \text{ a.s.} \quad (84)$$

Subtracting (84) from (83), we get

$$\lim_{m \rightarrow \infty} -\frac{1}{m} \log P(S^m | Y^m) = H_P(S_2 | Y_2 Y_1) \text{ a.s.} \quad (85)$$

Further, we have $\lim_{n \rightarrow \infty} \frac{M_n}{n} = 1-d$ almost surely. Using this with (85) in Lemma 2.4, we conclude that

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log P(S^{M_n} | Y^{M_n}) = (1-d)H_P(S_2 | Y_1 Y_2) \text{ a.s.} \quad (86)$$

We now argue that $-\frac{1}{n} \log P(S^{M_n} | Y^{M_n})$ is uniformly integrable. The $\text{Supp}(S^{M_n} | Y^{M_n})$ can be upper bounded by representing S^{M_n} as

$$\underbrace{xx \dots x}_{S_1} Y \underbrace{xx \dots x}_{S_2} Y \underbrace{x \dots x}_{S_3} \dots Y \underbrace{xx \dots x}_{S_{M_n}} Y \underbrace{xx \dots x}_{S_{M_n+1}}$$

where the Y 's represent the bits of the sequence Y^{M_n} , and each x represents a missing run. Since the maximum length of the above binary sequence is n , we have $\text{Supp}(S^{M_n} | Y^{M_n}) \leq 2^n$. Hence, from Lemma 2.2, $-\frac{1}{n} \log \Pr(S^{M_n} | Y^{M_n})$ is uniformly integrable.

Using this together with (86) in Lemma 2.3, we conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{n} H_P(S^{M_n} | Y^{M_n}) = \lim_{n \rightarrow \infty} \mathbf{E} \left[-\frac{1}{n} \log P(S^{M_n} | Y^{M_n}) \right] = \mathbf{E} \left[\lim_{n \rightarrow \infty} -\frac{1}{n} \log P(S^{M_n} | Y^{M_n}) \right] = (1-d)H_P(S_2 | Y_1 Y_2).$$

Thus we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} H_P(S^{M_n+1} | Y^{M_n}) = \lim_{n \rightarrow \infty} \frac{1}{n} H_P(S^{M_n} | Y^{M_n}) + \lim_{n \rightarrow \infty} \frac{1}{n} H_P(S_{M_n+1} | Y^{M_n}, S^{M_n}) = (1-d)H_P(S_2 | Y_1 Y_2) + 0.$$

□

B.2 Proof of Lemma 5.2

Due to (35), it is enough to show that $\frac{1}{n}H_P(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'})$ converges to $(1 - \gamma)H_P(L_1^X | L_1^{Y'})$. Since $\{(L_1^X, L_1^{Y'}), (L_2^X, L_2^{Y'}), \dots\}$ is an i.i.d process, from the strong law of large numbers, we have

$$\lim_{m \rightarrow \infty} -\frac{1}{m} \Pr(L_1^X, \dots, L_m^X | L_1^{Y'}, \dots, L_m^{Y'}) = H_P(L_1^X | L_1^{Y'}) \quad a.s. \quad (87)$$

Further, we have the normalized number of input runs $\frac{R_n}{n} \rightarrow (1 - \gamma)$ almost surely. Using the above in Lemma 2.4, we obtain

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'}) = H_P(L_1^X | L_1^{Y'}) \quad a.s. \quad (88)$$

We now argue that $-\frac{1}{n} \log \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'})$ is uniformly integrable. $\text{Supp}(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'})$ can be upper bounded by 2^n since the random sequence $(L_1^X, \dots, L_{R_n}^X)$ is equivalent to X^n , which can take on at most 2^n values. Hence, from Lemma 2.2, $-\frac{1}{n} \log \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'})$ is uniformly integrable. Using this together with (88) in Lemma 2.3, we conclude that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} H_P(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'}) &= \lim_{n \rightarrow \infty} \mathbf{E} \left[-\frac{1}{n} \log \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'}) \right] \\ &= \mathbf{E} \left[\lim_{n \rightarrow \infty} -\frac{1}{n} \log \Pr(L_1^X, \dots, L_{R_n}^X | L_1^{Y'}, \dots, L_{R_n}^{Y'}) \right] \\ &= H_P(L_1^X | L_1^{Y'}). \end{aligned} \quad (89)$$

□

C Deletion+Insertion Channel

C.1 Proof of Lemma 6.4

We have

$$\frac{1}{m} H(S^m | T^m, Y^m) = \sum_{j=1}^m \frac{1}{m} H(S_j | S^{j-1}, T^m, Y^m) \leq \sum_{j=1}^m \frac{1}{m} H(S_j | Y_{j-1}, Y_j, T_j)$$

We will show that $\lim_{j \rightarrow \infty} H(S_j | Y_{j-1}, Y_j, T_j)$ exists and obtain an analytical expression for it. For all j ,

$$\begin{aligned} H(S_j | Y_{j-1}, Y_j, T_j) &= P(Y_{j-1}, Y_j, T_j = 0) H(S_j | Y_{j-1}, Y_j, T_j = 0) = \\ &\sum_{y \in \{0,1\}} P(Y_{j-1} = Y_j = y, T_j = 0) H(S_j | Y_{j-1} = Y_j = y, T_j = 0) + P(Y_{j-1} = \bar{y}, Y_j = y, T_j = 0) H(S_j | Y_{j-1} = \bar{y}, Y_j = y, T_j = 0) \end{aligned} \quad (90)$$

The first equality above holds since $T_j = 1$ implies Y_j is an inserted bit, and so no deleted runs occur between Y_{j-1} and Y_j . $P(Y_{j-1}, Y_j, T_j = 0)$ can be computed as follows.

$$\begin{aligned}
P(Y_{j-1} = y, Y_j = y, T_j = 0) &= P(I_{j-1} = 0, Y_{j-1} = y, T_j = 0, Y_j = y) + P(I_{j-1} = 1, T_{j-1} = 0, Y_{j-1} = y, T_j = 0, Y_j = y) \\
&\quad + P(I_{j-1} = 1, T_{j-1} = 1, Y_{j-1} = y, T_j = 0, Y_j = y) \\
&\stackrel{(a)}{=} \frac{1}{2} P(I_{j-1} = 0) (P(I_j = 0 | I_{j-1} = 0) q + P(I_j = 1 | I_{j-1} = 0) \alpha) + \frac{1}{2} P(I_{j-1} = 1) \alpha q + \frac{1}{2} P(I_{j-1} = 1) \bar{\alpha} (1 - q) \\
&\xrightarrow{j \rightarrow \infty} \frac{1}{2} \left[\frac{1}{1+i'} ((1-i')q + i'\alpha) + \frac{i'}{1+i'} \alpha q + \frac{i'}{1+i'} \bar{\alpha} (1-q) \right].
\end{aligned} \tag{91}$$

The last two terms in (a) are obtained by noting that $I_{j-1} = 1$ implies Y_{j-1} is an insertion and hence $T_j = 0$. In this case, $Y_{j-2} = y$ corresponds to the last non-inserted bit before Y_j . The last line is due to the fact that $\{I_j\}_{j \geq 1}$ converges is a Markov chain that converges to the stationary distribution $P(I_j = 1) = \frac{i'}{1+i'}$, $P(I_j = 0) = \frac{1}{1+i'}$. Thus for sufficiently large j , $P(Y_{j-1} = y, Y_j = y, T_j = 0)$ is at most ϵ in total variation norm from the stationary distribution

$$\pi(Y_{j-1} = y, Y_j = y, T_j = 0) = \frac{1}{2} \left[\frac{1}{1+i'} ((1-i')q + i'\alpha) + \frac{i'}{1+i'} \alpha q + \frac{i'}{1+i'} \bar{\alpha} (1-q) \right], \quad y \in \{0, 1\}.$$

Similarly, $P(Y_{j-1} = y, Y_j = \bar{y}, T_j = 0)$ converges to

$$\begin{aligned}
\pi(Y_{j-1} = y, Y_j = \bar{y}, T_j = 0) &= \pi(I_{j-1} = 0, Y_{j-1} = y, T_j = 0, Y_j = \bar{y}) + P(I_{j-1} = 1, T_{j-1} = 0, Y_{j-1} = y, T_j = 0, Y_j = \bar{y}) \\
&\quad + \pi(I_{j-1} = 1, T_{j-1} = 1, Y_{j-1} = y, T_j = 0, Y_j = \bar{y}) \\
&\stackrel{(a)}{=} \frac{1}{2} \pi(I_{j-1} = 0) P(I_j = 0 | I_{j-1} = 0) (1-q) + \frac{1}{2} P(I_{j-1} = 1) \alpha (1-q) + \frac{1}{2} \pi(I_{j-1} = 1) \bar{\alpha} q \\
&= \frac{1}{2} \left[\frac{1}{1+i'} (1-i') (1-q) + \frac{i'}{1+i'} \alpha (1-q) + \frac{i'}{1+i'} \bar{\alpha} q \right].
\end{aligned} \tag{92}$$

We next determine the joint distributions $\pi(S_j, Y_{j-1} = \bar{y}, Y_j = y, T_j = 0)$ and $\pi(S_j, Y_{j-1} = \bar{y}, Y_j = y, T_j = 0)$ to compute $H_\pi(S_j | Y_{j-1} = \bar{y}, Y_j = y, T_j = 0)$ and $H_\pi(S_j | Y_{j-1} = y, Y_j = y, T_j = 0)$ in (90). For $k = 0, 1, \dots$, we have

$$\begin{aligned}
\pi(S_j = k, Y_{j-1} = y, Y_j = y, T_j = 0) &= \\
&\pi(I_{j-1} = 0, Y_{j-1} = y, T_j = 0, Y_j = y, S_j = k) + \pi(I_{j-1} = 1, T_{j-1} = 0, Y_{j-1} = y, T_{j=0}, Y_j = y, S_j = k) \\
&\quad + \pi(I_{j-1} = 1, T_{j-1} = 1, Y_{j-1} = y, T_{j=0}, Y_j = y, S_j = k).
\end{aligned} \tag{93}$$

The first term corresponds to Y_{j-1} being an original input bit, the second term to Y_{j-1} being a duplication, and the third to Y_{j-1} being a complementary insertion, respectively. Each of these terms can be calculated in

a manner very similar to equations (29) and (30) of Proposition 5.2. Hence we obtain,

$$\pi(S_j = k, Y_{j-1} = y, Y_j = y, T_j = 0) = \begin{cases} \frac{1}{2(1+i')} \left[i'\alpha + (1 - i'\bar{\alpha}) \frac{\gamma(1-d)}{1-\gamma d} + i'\bar{\alpha} \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2} \right], & k = 0 \\ \frac{1-i'\bar{\alpha}}{2(1+i')} \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2} \left(\frac{d(1-\gamma)}{1-\gamma d} \right)^k, & k = 1, 3, \dots \\ \frac{i'\bar{\alpha}}{2(1+i')} \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2} \left(\frac{d(1-\gamma)}{1-\gamma d} \right)^k, & k = 2, 4, \dots \end{cases} \quad (94)$$

Similarly, we also determine

$$\pi(S_j = k, Y_{j-1} = \bar{y}, Y_j = y, T_j = 0) = \begin{cases} \frac{1}{2(1+i')} \left[(1 - i'\bar{\alpha}) \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2} + i'\bar{\alpha} \frac{\gamma(1-d)}{1-\gamma d} \right], & k = 0 \\ \frac{i'\bar{\alpha}}{2(1+i')} \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2} \left(\frac{d(1-\gamma)}{1-\gamma d} \right)^k, & k = 1, 3, \dots \\ \frac{1-i'\bar{\alpha}}{2(1+i')} \frac{(1-\gamma)(1-d)}{(1-\gamma d)^2} \left(\frac{d(1-\gamma)}{1-\gamma d} \right)^k, & k = 2, 4, \dots \end{cases} \quad (95)$$

From (91) and (94), we can compute

$$\begin{aligned} & \pi(Y_{j-1} = y, Y_j = y, T_j = 0) H_\pi(S_j | Y_{j-1} = y, Y_j = y, T_j = 0) \\ &= \frac{1}{2(1+i')} \left[(i'\alpha + (1 - i'\bar{\alpha}) \frac{\gamma(1-d)}{1-\gamma d} + i'\bar{\alpha} \beta) \log_2 \left(\frac{i'\alpha + (1 - i'\bar{\alpha})q + i'\bar{\alpha}\bar{q}}{i'\alpha + (1 - i'\bar{\alpha}) \frac{\gamma(1-d)}{1-\gamma d} + i'\bar{\alpha}\beta} \right) + i'\bar{\alpha}\beta \frac{2\theta^2}{(1-\theta^2)^2} \log_2 \left(\frac{1}{\theta} \right) \right. \\ & \quad \left. + (1 - i'\bar{\alpha})\beta \frac{\theta(1+\theta^2)}{(1-\theta^2)^2} \log_2 \left(\frac{1}{\theta} \right) + \frac{\theta\beta(1-i'\bar{\alpha})}{1-\theta^2} \log_2 \left(\frac{i'\alpha + (1 - i'\bar{\alpha})q + i'\bar{\alpha}\bar{q}}{\beta(1-i'\bar{\alpha})} \right) + \frac{\theta^2\beta i'\bar{\alpha}}{1-\theta^2} \log_2 \left(\frac{i'\alpha + (1 - i'\bar{\alpha})q + i'\bar{\alpha}\bar{q}}{\beta i'\bar{\alpha}} \right) \right] \end{aligned} \quad (96)$$

where θ and β are defined in the statement of the lemma. Similarly, from (92) and (95), one can compute

$$\begin{aligned} & \pi(Y_{j-1} = y, Y_j = y, T_j = 0) H_\pi(S_j | Y_{j-1} = y, Y_j = y, T_j = 0) \\ &= \frac{1}{2(1+i')} \left[(i'\bar{\alpha} \frac{\gamma(1-d)}{1-\gamma d} + (1 - i'\bar{\alpha})\beta) \log_2 \left(\frac{(1 - i'\bar{\alpha})\bar{q} + i'\bar{\alpha}q}{i'\bar{\alpha} \frac{\gamma(1-d)}{1-\gamma d} + (1 - i'\bar{\alpha})\beta} \right) + i'\bar{\alpha}\beta \frac{\theta(1+\theta^2)}{(1-\theta^2)^2} \log_2 \left(\frac{1}{\theta} \right) \right. \\ & \quad \left. + (1 - i'\bar{\alpha})\beta \frac{2\theta^2}{(1-\theta^2)^2} \log_2 \left(\frac{1}{\theta} \right) + \frac{\theta^2\beta(1-i'\bar{\alpha})}{1-\theta^2} \log_2 \left(\frac{(1 - i'\bar{\alpha})\bar{q} + i'\bar{\alpha}q}{\beta(1-i'\bar{\alpha})} \right) + \frac{\theta\beta i'\bar{\alpha}}{1-\theta^2} \log_2 \left(\frac{(1 - i'\bar{\alpha})\bar{q} + i'\bar{\alpha}q}{\beta i'\bar{\alpha}} \right) \right]. \end{aligned} \quad (97)$$

Substituting (96) and (97) in (90) completes the proof of the lemma. \square